

Západočeská univerzita v Plzni
Fakulta aplikovaných věd
Katedra informatiky a výpočetní techniky

Diplomová práce

Neparametrické jádrové odhady a jejich praktická aplikace

Prohlášení

Prohlašuji, že jsem diplomovou práci vypracoval samostatně a výhradně s použitím citovaných pramenů.

V Plzni dne

Podpis:
Pavel Brom

Poděkování

Děkuji panu Ing. Tomáši Ťoupalovi, Ph. D. za cenné rady, připomínky, ochotu a čas věnovaný vedení této diplomové práce. Rovněž děkuji své rodině a přítelkyni za podporu a trpělivost během celého studia.

Neparametrické jádrové odhady a jejich praktická aplikace

Abstrakt

Diplomová práce se zabývá problematikou neparametrických jádrových odhadů, především regresní křivky. V úvodní části je provedena rešerše na toto téma, ve které je popisován parametrický přístup k odhadu funkce hustoty a regresní křivky a následně vybrané neparametrické metody vedoucí k jádrovým odhadům. Za oblast jejich praktické aplikace bylo zvoleno obchodování s akciemi na burze cenných papírů na českém trhu. Následně je proto popsán způsob, jakým se na burze obchoduje, a náklady s tím spojené. Jádrové odhady jsou v této práci srovnávány s klouzavými průměry, které jsou k obchodování využívány v technické analýze akcií. Pro každý vybraný akciový titul zvlášť je optimalizován model i parametry pomocí simulací provedených na základě reálných dat za minimálně čtyři roky. Toto nastavení je nakonec ověřováno navrženým obchodním systémem na reálných datech v rámci více než dvouletého období.

Klíčová slova

regresní křivka, klouzavé průměry, jádrový odhad, Nadaraya-Watson, obchodování s akciemi, technické indikátory, obchodní systém

Nonparametric kernel estimation and practical application

Abstract

This master's thesis deals with nonparametric kernel estimates of a regression function. The summary of this topic is provided in the first part of the thesis. Firstly, the parametric approach to the estimation of a density and regression functions is mentioned. Subsequently, the thesis focuses on selected nonparametric methods, most importantly on kernel estimates. Stock trading in the Czech market was chosen as the area of their practical application. Consequently, the way how trading on a stock-exchange works and transaction fees are discussed. Kernel estimates are compared to the moving averages used in the stock technical analysis. The proposed models and their parameters are optimized in simulations based on at least five-year long real data for each selected stock. Afterwards, this tailor-made setting is used in a simulation of a designed trading system, which is based on more than two-year long real data.

Keywords

Regression function, moving averages, kernel estimate, Nadaraya-Watson, stock trading, technical indicators, trading system

Obsah

1	Úvod	1
2	Problematika jádrových odhadů	2
2.1	Odhad neznámé hustoty pravděpodobnosti	3
2.1.1	Histogram	3
2.1.2	„Jádrový“ přístup.....	5
2.1.2.1	Druhy jádrových funkcí	5
2.1.2.2	Jádrový odhad funkce hustoty	6
2.1.2.3	Volba vyhlazovacího parametru pro jádrový odhad hustoty.....	7
2.1.3	Srovnání parametrického a neparametrického odhadu	9
2.2	Odhad neznámé distribuční funkce	10
2.3	Odhad neznámé regresní křivky.....	10
2.3.1	Regrese.....	12
2.3.2	Regresogram	13
2.3.3	Klouzavé průměry	14
2.3.4	„Jádrový“ přístup.....	15
2.3.4.1	Nadaraya-Watsonův jádrový odhad regresní křivky.....	16
2.3.4.2	Rozbor vlivu parametru vyhlazení na výsledný odhad.....	20
2.3.4.3	Volba vyhlazovacího parametru pro jádrový odhad regresní křivky	21
3	Obchodování s akciemi na BCPP	24
3.1	Poplatky spojené s obchodováním na BCPP	25
3.2	Daň z příjmu	25
4	Technická analýza akcií	27
4.1	Definice technické analýzy.....	27
4.2	Studování grafů	28
4.3	Technické indikátory	28
4.3.1	Klouzavé průměry	29
4.3.1.1	Jednoduchý klouzavý průměr	29
4.3.1.2	Vážený klouzavý průměr	30
4.3.1.3	Exponenciální klouzavý průměr	31
4.3.1.4	Klouzavá regrese	32
4.3.1.5	Použití více klouzavých průměrů zároveň.....	33
4.3.1.6	Možná strategie obchodování	36
4.3.2	Další indikátory.....	36

5	Data a jejich úprava.....	37
5.1	Vliv dividendy na kurz a jeho eliminace	37
5.2	Vliv štěpení akcií na kurz a jeho eliminace.....	38
5.3	Data	38
5.3.1	ČEZ, a.s.	39
5.3.2	Komerční banka, a.s.	39
5.3.3	Philip Morris ČR, a.s.	40
6	Obchodní systém.....	42
6.1	Popis navržených modelů	42
6.1.1	Modely založené na klouzavých průměrech.....	42
6.1.2	Modely založené na jádrových odhadech.....	42
6.1.2.1	Použití jádrového odhadu	42
6.1.2.2	„Zpožděný“ jádrový odhad.....	43
6.1.2.3	Odhad vyhlazovacího parametru a parametru posunu	44
6.1.2.4	Jednotlivé modely	45
6.2	Optimalizace parametrů a modelu	45
6.2.1	Princip „optimalizace“	45
6.2.2	Optimální nastavení pro jednotlivé akciové tituly	49
6.2.3	Popis optimalizačního algoritmu.....	50
6.3	Simulace obchodování	52
6.3.1	Předpoklady pro obchodování.....	52
6.3.2	Popis algoritmu	52
6.3.3	Akcie ČEZ.....	53
6.3.4	Akcie KB.....	54
6.3.5	Akcie PM.....	55
6.4	Modifikace optimalizačního procesu	56
6.4.1	Optimalizace parametrů a modelů	57
6.4.2	Simulace obchodování	58
6.4.2.1	Akcie ČEZ.....	58
6.4.2.2	Akcie KB.....	58
6.4.2.3	Akcie PM.....	59
7	Závěr.....	61
	Seznam použitých zkratk a značení.....	63
	Seznam obrázků	65
	Seznam tabulek.....	67

Reference	68
Příloha A – přiložené soubory	73
Příloha B – zdrojové kódy.....	74

1 Úvod

Neparametrickým odhadům se v posledních letech díky rozvoji výpočetní techniky dostává více pozornosti. Počítače a software poskytují nástroje ke stále přesnějším a rychlejším výpočtům. Jedná se o poměrně mladou disciplínu matematické statistiky, jejíž první odhady byly podle [1], str. 150, „navrženy počátkem 50. let pro odhad spektrální hustoty a od té doby byly velmi zdokonaleny“ a oblast jejich použití se rozšiřovala. Problematika jádrového odhadování je velice rozsáhlá. Konkrétně jádrové odhady jsou považovány za efektivní prostředky neparametrického odhadování.

V této práci je pozornost věnována odhadům regresní křivky využívajícím odhady funkce hustoty. Obě tyto oblasti jsou proto blíže analyzovány v kapitole 2. V obou případech je nejprve popsán parametrický přístup k danému problému a porovnán s vybranými neparametrickými metodami, včetně jádrových odhadů.

Ve třetí kapitole je uvedena aplikační sféra, pro kterou jsou následně navrhovány a testovány modely založené na jádrových odhadech. Jedná se o obchodování s akciemi vybraných českých akciových společností na české burze cenných papírů. V této kapitole je především popsán způsob, jak lze s akciemi na burze obchodovat a jaké náklady jsou s tím spojeny.

Čtvrtá kapitola je zaměřena na technickou analýzu akcií. Jde o disciplínu zabývající se obchodováním s akciemi a analýzami, jak maximalizovat zisk z investice do akcií. Především jsou v této kapitole představeny různé druhy klouzavých průměrů, které techničtí analytici využívají. Klouzavé průměry i jádrové odhady přikládají každému pozorování určitou „váhu“. Na základě tohoto společného znaku mohou být oba přístupy dále porovnávány.

V páté kapitole je řešen vliv dividendy a štěpení akcií na kurz. Také jsou stručně popsány akciové tituly, pro které byly získány reálné historické kurzy, na kterých jsou navržené modely testovány. Veškeré výpočty v praktické části práce jsou založeny na reálných historických závěrečných kurzech akcií očištěných způsobem uvedeným v této kapitole.

V závěrečné šesté kapitole jsou navrženy modely obchodního systému, které jsou postaveny na propojení poznatků ze druhé a čtvrté kapitoly. Pozornost je věnována odvození „zpožděných“ modelů využívajících jádrové odhady. Následuje numerická optimalizace modelů a parametrů pro jednotlivé akciové tituly, jejichž výstupem jsou modely a kombinace parametrů, které jsou schopny „nejlépe“ (ve smyslu maximalizace mediánu zisku z obchodů) identifikovat okamžiky pro nákup a prodej akcií. Vždy je tak dosaženo nastavení obchodního systému „na míru“ akciovému titulu. Ve třetí části kapitoly je vytvořen obchodní systém používající „optimální“ modely a parametry.

V této práci jsou použity obrázky vytvořené vlastními skripty. Algoritmy vytvořené pro navržený proces optimalizace a obchodní systém jsou popsány v odpovídajících částech šesté kapitoly. Všechny tyto skripty jsou přiloženy na CD a jejich seznam s krátkými popisky je k dispozici v Příloha A – přiložené soubory. Zdrojové kódy důležitých skriptů jsou v plném znění k dispozici v Příloha B – zdrojové kódy. Veškeré výpočty a simulace založené na výše zmíněných algoritmech jsou provedeny v software MATLAB R2015a. Součástí práce je také výstup optimalizačního skriptu v podobě souboru, který byl následně doupřeván a zřehledněn v software Microsoft Excel 2013.

2 Problematika jádrových odhadů

Tato část vychází z [1], [2], [3], [4].

V situaci, kdy jsou k dispozici data, o kterých není známo nic konkrétního, a je třeba zjistit, jaké jsou jejich vlastnosti a případně předpovědět budoucí vývoj, je nutné najít model, který nejlépe (např. ve smyslu minimalizace určitého kritéria) vystihuje jejich charakter. Tento model je nazýván regresní křivka. Proces vedoucí k nalezení odhadu regresní křivky je nazýván vyhlazování. K vyhlazování lze využít dva přístupy:

- parametrické odhady – už v prvním kroku je nezbytné vhodně zvolit funkci (resp. model) závisující na parametrech, které jsou následně odhadovány, a
- neparametrické odhady – není k dispozici žádná funkce (ani její odhad) a analyzována jsou samotná data.

Parametrické odhady regresních modelů je možné dále rozdělit do dvou skupin:

- lineární regrese – modely lineární v parametrech, např. přímka $y = a + bx$ a
- nelineární regrese – modely nelineární v parametrech, např. exponenciála $y = ab^x$ (v některých případech je možné pomocí lineární transformace převést na lineární regresi).

Tato práce se vzhledem k zadání bude zabývat především neparametrickými odhady, konkrétně jádrovými odhady, a jejich aplikací. Ty se používají k odhadům různých funkcí, např. funkce hustoty, distribuční funkce nebo regresní křivky, přičemž zde budou primárně použity k aproximaci regresní křivky. Platí zde omezení na případy jednorozměrných náhodných veličin. V případě neparametrických odhadů nejsou uvažovány modely ve smyslu regrese, tj. není k dispozici žádný předpis funkce, která by modelovala dané hodnoty, ani přibližný model. V tomto případě jde o metody založené jen na naměřených pozorováních a aproximacích tvaru neznámé „původní“ funkce.

Výhodou neparametrických odhadů je, že k provedení odhadu není potřeba žádná apriorní informace¹ o datech ani způsobu jejich sběru (v praxi taková informace vůbec nemusí být k dispozici) – tyto odhady tedy mohou být použity pro odhad libovolného rozdělení. Najít „běžné“ rozdělení (např. normální, ...), ze kterého by mohla výběrová data pocházet, navíc bývá složité a někdy nemožné. Nevýhodou je, že se vychází pouze ze získaných dat a nezbyvá než důvěřovat jejich původu, tj. správnost měření, apod.

Jádrové odhady jsou používány k aproximaci neznámé funkce pomocí vážených průměrů² jednotlivých pozorování. Pro odhad funkce v konkrétním bodě jsou odpovídající váhy přiřazeny těm pozorováním, která náležejí do některého symetrického okolí odhadovaného bodu (podrobněji je vysvětleno v částech 2.1.2.1 a 2.1.2.2, ve kterých jsou pro pochopení popisované problematiky k dispozici Obrázek 2.3 a Obrázek 2.4). Jak je dále ukázáno, kvalita jádrových odhadů závisí především na „vhodné“ volbě vyhlazovacího parametru (tj. šířky vyhlazovacího okna). Volba jádrové

¹ Apriorní informace je informace, kterou máme předem (apriori) k dispozici. (viz [22], str. 55)

² Váhy jsou určeny zvolenou jádrovou funkcí.

funkce není považována za příliš podstatnou v porovnání s volbou parametru vyhlazení. Nicméně toto tvrzení nemusí být obecně platné, jak je ukázáno v praktické části práce, např. v Tabulka 3 v části 6.2.1.

2.1 Odhad neznámé hustoty pravděpodobnosti

Nechť je k dispozici výběr ze spojitě náhodné veličiny X , tj. x_1, x_2, \dots, x_n , kde n je pevný počet pozorování, a je odhadována funkce hustoty³, tj. $f(x)$.

2.1.1 Histogram

Tato část vychází z [1], [5] a [6].

Jednou z použitelných metod odhadu neznámé hustoty pravděpodobnosti je sestavení histogramu. To, jak se uvádí v [5], str. 6, je považováno za jeden z nejstarších přístupů. Jeho nevýhodou však je, že výsledný odhad je po částech spojitá funkce, přestože je předpokládána spojitá funkce hustoty, a že se jedná o hrubý odhad. Dále je za nevýhody histogramu považována jeho citlivost na velikost a počet tříd a složitost přechodu k vícerozměrnému rozdělení. Jedná se ale o užitečnou metodu, díky které se získá prvotní představa o tvaru skutečné hustoty, o šikmosti⁴ a špičatosti⁵ dat, apod.

Histogram je grafický způsob zobrazení tvaru pravděpodobnostního rozdělení. Jedná se o sloupcový graf, ve kterém je osa x rozdělena do k tříd (podle [6], str. 35, můžeme určit optimální⁶ počet tříd⁷ např. pomocí Sturgesova pravidla: $k \approx 1 + 3.32 \log n$ nebo Yuleova pravidla: $k \approx 2.5 \sqrt[4]{n}$) a osa y zachycuje četnosti⁸ dat v jednotlivých třídách, tj. ta pozorování x_i , která leží ve vybrané třídě spolu s x , např. $x \pm \frac{h}{2}$, kde h je šířka třídy. Odhad hustoty označovaný $\hat{f}(x)$ lze pro uvedenou situaci zapsat ve tvaru funkce

$$\hat{f}(x) = \frac{1}{nh} \sum_{i=1}^n I_{\left[-\frac{1}{2}; \frac{1}{2}\right]} \left(\frac{x_i - x}{h} \right), \quad (1)$$

kde $h \in \mathbb{R}^+$ je šířka intervalu,

$x_i \in \mathbb{R}$ je i -té pozorování náhodné veličiny X a

$I_{[a;b]}(x) = \begin{cases} 1 & a \leq x \leq b; a, b \in \mathbb{R}, \\ 0 & \text{jinak,} \end{cases}$ je tzv. indikátorová funkce.

Příklad histogramu vytvořeného pomocí Sturgesova pravidla ukazuje Obrázek 2.1.

³ Funkce hustoty říká, jak jsou data „rozdělena“. Díky znalosti $f(x)$ je známa také charakteristika polohy (např. střední hodnotu) a charakteristika rozptýlení (např. rozptyl), pokud existují, a dále pravděpodobnost, s jakou hodnoty budou v určitém intervalu. (viz [3], str. 21)

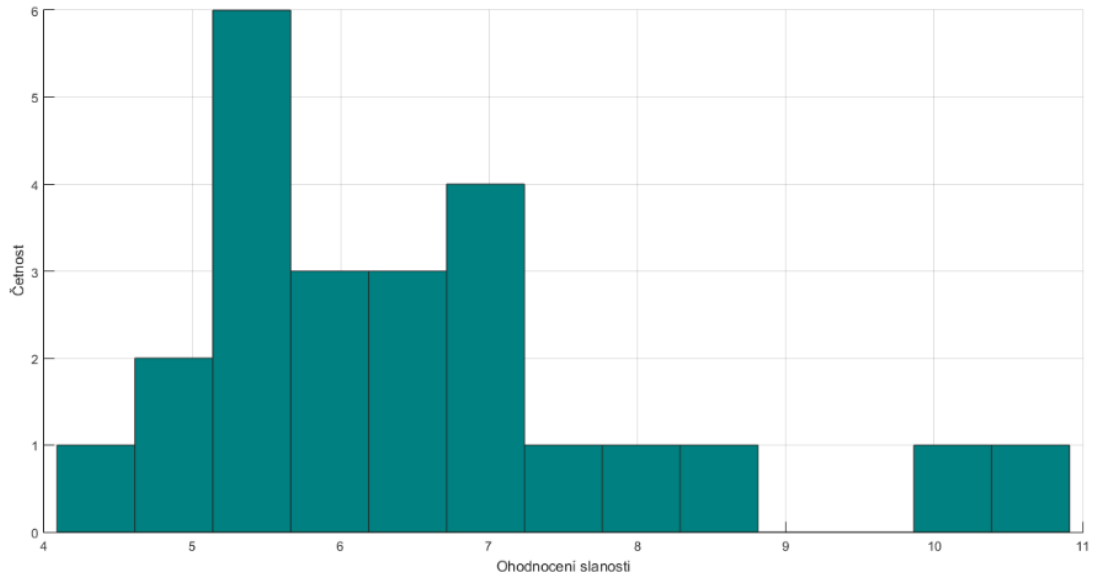
⁴ „Šikmost měří nesouměrnost rozložení četností kolem průměru.“ [62], str. 12

⁵ „Špičatost měří koncentraci rozložení četností kolem průměru.“ [62], str. 12.

⁶ k se zaokrouhluje podle uvážení. Dále podle [6], str. 35, „počátek prvního intervalu volíme tak, aby nejmenší a největší hodnota padla do prvního a posledního intervalu“. Při konstrukci histogramu proto může dojít k navýšení počtu tříd. Výpočet „optimálního“ k je tedy orientační.

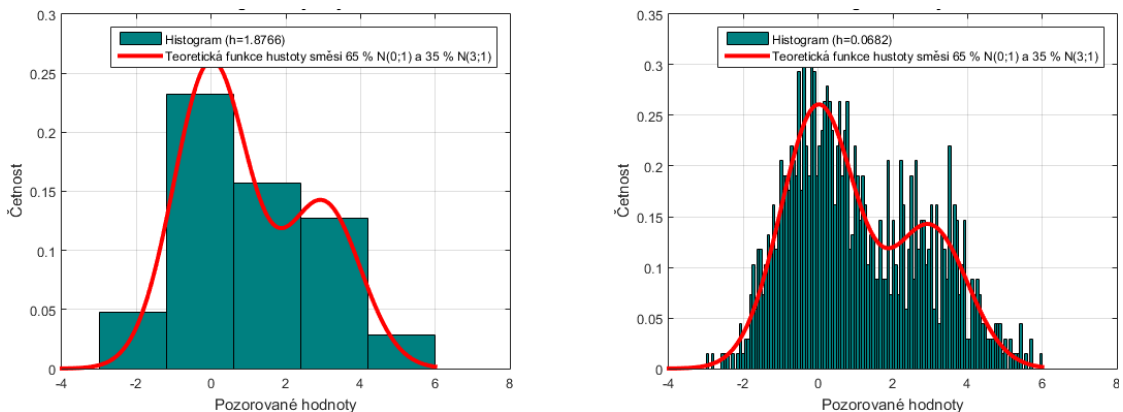
⁷ V této práci jsou uvažovány ekvidistantní (tj. stejně široké) třídy.

⁸ V této práci jsou uvažovány absolutní četnosti (blíže viz [6], str. 29 a 33).



Obrázek 2.1: Příklad histogramu pro hodnocení slanosti 16% krémového sýru, pro výpočet počtu tříd použito Sturgesovo pravidlo, $h \doteq 0.5250$, $n = 24$ (zdroj: vlastní zpracování⁹)

Vyhazení odhadu $\hat{f}(x)$ závisí pouze na h , tj. šířce tříd. O h lze hovořit jako o parametru této metody. Třídy nesmí být „příliš“ široké (splnutí tříd) ani „příliš“ úzké (signalizace „falešných rysů“ rozdělení nebo až „rozpadnutí“ histogramu na více částí, přestože funkce hustoty nemusí mít více „vrcholů“ nebo být multimodální¹⁰). V obou případech může dojít ke zkreslení tvaru hustoty a odhad povede ke zkresleným výsledkům při dalším zpracování. Situaci vysvětluje Obrázek 2.2 vygenerovaný ze směsi¹¹ dvou normálních rozdělení (65 % $N(0;1)$ a 35 % $N(3;1)$) pro 1 000 hodnot. Proto je pro tento úkol zásadní hledání „vhodného“ h . S tímto problémem, jak už bylo naznačeno, dokáže pomoci mj. zmiňované Sturgesovo nebo Yuleovo pravidlo. Ze znalosti počtu tříd lze vypočítat šířku intervalu pomocí vztahu $h = \frac{x_{max} - x_{min}}{k}$ (viz [6], str. 35). Pro konstruování histogramů je podle [7], str. 8, nejčastěji používaným pravidlem Sturgesovo pravidlo.



Obrázek 2.2: Srovnání histogramů s různými počty (resp. šířkami – vepsané v legendách u obrázků) tříd a známé hustoty, $n = 1\,000$ (zdroj: vlastní simulace)

⁹ Data pocházejí ze serveru *Plant Food Science group & Spectroscopy and Chemometrics* [55] – konkrétně jde o studii zaměřenou na statistické zpracování hodnocení vybraných vlastností neoznačených krémových sýrů.

¹⁰ Funkce hustoty multimodálního rozdělení má více modů (viz [57]). Modus je hodnota ze souboru, která se v něm vyskytuje nejčastěji (viz [18], str. 18.).

¹¹ Konečná směs rozdělení o g komponentách se definuje pomocí směsi hustot $f_i: f(y_j) = \sum_{i=1}^g \pi_i f_i(y_j)$, kde f_i je funkce hustoty i -té komponenty a $\pi_i \in (0, 1)$ je váha i -té komponenty a platí $\sum_{i=1}^g \pi_i = 1$. (blíže viz [63], str. 6)

Druhým přístupem je přímo odhad šířky intervalu. Podle [8], str. 59, k tomu lze použít např. Scottovo pravidlo¹²: $h = \frac{3.5 s_X}{\sqrt[3]{n}}$, kde s_X je výběrová směrodatná odchylka¹³ pro náhodnou veličinu X , nebo robustní¹⁴ Freedmanovo a Diaconisovo pravidlo: $h = \frac{2 \cdot IQR}{3\sqrt[3]{n}}$, kde $IQR = \frac{x_{0.75} - x_{0.25}}{2}$ je kvartilová odchylka¹⁵.

Přestože jsou k dispozici pravidla pro určení počtu i šířky intervalu, nejsou tyto „parametry“ dány jednoznačně. Vždy záleží na subjektivním posouzení konkrétní situace. Intervaly by podle [6], str. 36, měly být zkonstruovány tak, „*aby odpovídající rozdělení četností dokázalo vypovídat alespoň o základních vlastnostech sledovaného znaku*“.

2.1.2 „Jádrový“ přístup

V této části je ukázáno, jaké vlastnosti musí funkce mít, aby mohla být jádrovou funkcí, dále jsou představeny nejpoužívanější jádrové funkce. Nakonec je uvedena metoda odhadu vyhlazovacího parametru.

2.1.2.1 Druhy jádrových funkcí

Jádrová funkce $k(x)$ může být jakákoliv spojitá funkce, která podle [9], str. 2, splňuje následující podmínky¹⁶:

1. $k(x) = k(-x)$, tzn., že jádrová funkce je sudá, symetrická funkce,
2. $\int_{-\infty}^{+\infty} k(x) dx = 1$, tzn., že jádrová funkce může být pouze funkce hustoty,
3. $\int_{-\infty}^{+\infty} xk(x) dx = 0$, tzn., že jádrová funkce je centrovaná a (2)
4. $\int_{-\infty}^{+\infty} x^2k(x) dx = 1$, tzn., že jádrová funkce je normovaná¹⁷.

V Tabulka 1 jsou uvedeny příklady jádrových funkcí, které splňují výše uvedené předpoklady. Poslední sloupec tabulky udává, pro která x uvedený předpis funkce platí. Všude jinde (u všech $k(x)$ kromě Gaussovy jádrové funkce) platí $k(x) = 0$. Nejedná se o jediné možné jádrové funkce (pro některé další viz např. [1], str. 151), v tabulce jsou představeny některé používané¹⁸ a „běžné“ jádrové funkce podle [10], str. 23.

¹² Za předpokladu, že výběr pochází ze spojitého rozdělení (blíže viz [7], str. 1).

¹³ Podle [15], str. 92, měří směrodatná odchylka variabilitu dat v původních jednotkách. V případě spojitě náhodné veličiny s funkcí hustoty $f(x)$ pro její výpočet platí: $s_X = \sqrt{\int_{-\infty}^{\infty} x^2 f(x) dx - (\int_{-\infty}^{\infty} x f(x) dx)^2}$.

¹⁴ Robustní odhad je takový odhad, u kterého nemají vychýlené hodnoty na vstupu (způsobené např. chybou měření) „příliš“ velký vliv na samotný odhad. (blíže viz [56], str. 2)

¹⁵ „Hodnota kvartilové odchylky udává průměrnou vzdálenost mezi dvěma jednotlivými sousedními kvartily.“ [6], str. 46-47.

¹⁶ Běžně se pracuje s jádrovými funkcemi druhého řádu (viz [9], str. 3) (řád jádrové funkce „odpovídá předpokládanému počtu derivací neznámé hustoty“ [2], str. 29). Proto jsou podmínky 3 a 4 v (2) podle toho rovnou upraveny.

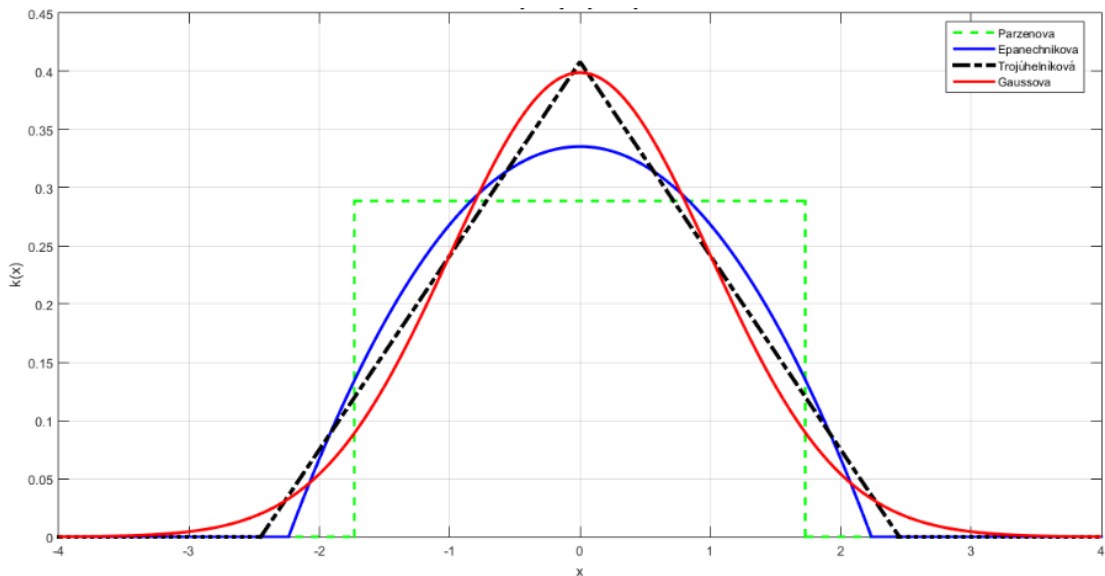
¹⁷ Tato podmínka byla oproti předloze upravena – místo nenulovosti daného integrálu, je požadována přímo hodnota 1, aby byla jádrová funkce normovaná (za platnosti třetí podmínky).

¹⁸ Např. také program MATLAB nabízí pod příkazem „fitdist“ právě tyto vybrané jádrové funkce k provedení jádrového odhadu.

Název jádrové funkce	Předpis jádrové funkce $k(x)$	Omezení
Parzenova (obdélníková, rovnoměrná)	$\frac{1}{2\sqrt{3}}$	$-\sqrt{3} \leq x \leq \sqrt{3}$
Epanechnikova	$\frac{3}{4\sqrt{5}} \left(1 - \left(\frac{x}{\sqrt{5}}\right)^2\right)$	$-\sqrt{5} \leq x \leq \sqrt{5}$
Trojúhelníková	$\frac{1}{\sqrt{6}} \left(1 - \frac{ x }{\sqrt{6}}\right)$	$-\sqrt{6} \leq x \leq \sqrt{6}$
Gaussova	$\frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$	$-\infty \leq x \leq \infty$

Tabulka 1: Vybrané jádrové funkce hustoty (pro výpočty viz [10], str. 38-39)

Podle [10], str. 24, nebo [1], str. 151-152, je v literatuře nejčastěji zmiňováno a pracováno s Gaussovou jádrovou funkcí pro její vlastnosti. Navíc je obsažena v téměř každém software, kde je možné použít statistické metody. Často se také doporučuje používat Epanechnikovu jádrovou funkci pro její jednoduché použití. Tvary jádrových funkcí z Tabulka 1 jsou zachyceny na Obrázek 2.3.


 Obrázek 2.3 Graficky zachycené průběhy vybraných jádrových funkcí $k(x)$ (zdroj: vlastní simulace)

2.1.2.2 Jádrový odhad funkce hustoty

Jádrové odhady velice často nacházejí své uplatnění právě při odhadech neznámé hustoty pravděpodobnosti. Rosenblattův obecný jádrový odhad podle¹⁹ [5], str. 9, má tvar

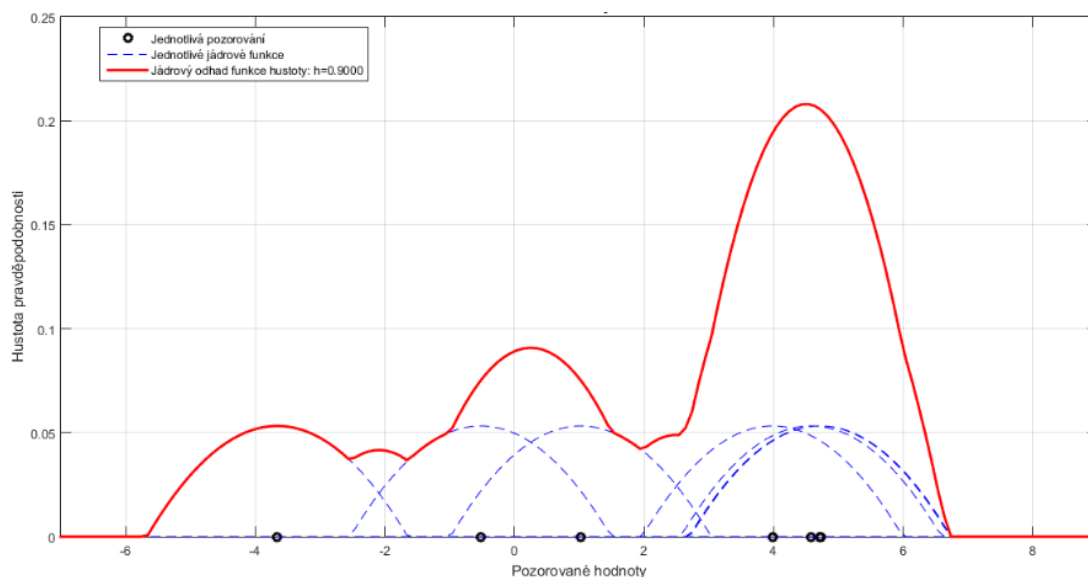
$$\hat{f}_h(x) = \frac{1}{nh} \sum_{i=1}^n k\left(\frac{x - x_i}{h}\right), \quad (3)$$

kde $h \in \mathbb{R}^+$ je parametr měřítka (nebo také vyhlazovací parametr, šířka okna),
 $k(x)$ je jádrová funkce aproximace splňující podmínky (2) a
 $x_i \in \mathbb{R}$ je i -té pozorování náhodné veličiny X .

¹⁹ V předpisu funkce bylo pouze pozměněno značení pro jádrovou funkci (z $K(\bullet)$ na $k(\bullet)$) z důvodu přehledného značení v další části textu.

Pokud je vyhlazovací parametr h na něčem závislý, pak pouze nanejvýš na rozsahu výběru n nebo na pozorovaných hodnotách x_i a platí: $\lim_{n \rightarrow +\infty} h = 0$.

Jak je uvedeno v [2], str. 29, jádrový odhad se získá tak, že v každém bodě x_i je sestrojena jádrová funkce k a odhad v bodě x je průměr funkčních hodnot n jádrových funkcí v tomto bodě. Konstrukce je naznačena na Obrázek 2.4, kde jsou čárkovaně vykresleny jednotlivé jádrové funkce (konkrétně Epanechnikovy) v bodech x_i a plnou čarou odhad funkce hustoty. Demonstrace konstrukce jádrového odhadu funkce hustoty pomocí (3) je zachycena na Obrázek 2.6.



Obrázek 2.4: Příklad konstrukce jádrového odhadu hustoty za použití Epanechnikovy jádrové funkce, $h = 0.9$, $n = 7$ (zdroj: vlastní simulace)

V [9], str. 3, jsou uvedeny výsledky studie použití jádrových odhadů v oblasti analýzy dat. Podle této studie je pro výsledný odhad funkce hustoty mnohem důležitější volba hodnoty vyhlazovacího parametru h než volba jádrové funkce $k(x)$. Nízké hodnoty h způsobují „kotrbatý“ odhad s velkým rozptylem, který vykazuje „falešné“ rysy. Naopak vysoké hodnoty h vedou k „příliš“ hladkému odhadu s malým rozptylem. Ani v jednom případě odhad nezachytí strukturu hustoty rozdělení. Výsledný odhad za použití nízkého h bývá (viz [11], str. 44) nazýván „podhlazený“, v opačném případě „přehladený“.

Volba (resp. změna) jádrové funkce (v rámci stejného řádu²⁰ jádrové funkce) ovlivňuje podle [8], str. 145, kvalitu výsledného odhadu hustoty velice slabě. V [8], str. 151, je dále ukázáno, že Epanechnikovu jádrovou funkci lze považovat za „optimální“²¹, proto je v obrázcích používána jako vzorová.

2.1.2.3 Volba vyhlazovacího parametru pro jádrový odhad hustoty

Tato část vychází z [10], [12] a [13].

Pro kvantifikaci parametru h je nutné zvolit nějaké kritérium, na základě kterého je následně možné určit „vhodnou“ hodnotu vyhlazovacího parametru tak, aby nedošlo k podhlazení nebo přehladení odhadu. V literatuře (např. [12], str. 27-52) je možné najít několik metod výpočtu parametru h ,

²⁰ Viz poznámka pod čarou č. 16.

²¹ ve smyslu nejhladší hustoty při „nevhodné“ volbě parametru měřítka h

které lze rozdělit do dvou kategorií podle toho, jestli je skutečná funkce hustoty známá. Proces optimalizace h je uvažován v pozorovaných hodnotách $x_i, i = 1, \dots, n$.

Jelikož analýza volby parametru vyhlazení není předmětem této práce, jsou v této části uvedeny vybrané metody, které byly vybrány jako zástupci ve své kategorii z důvodů uvedených níže. Tato problematika je podrobněji popsána např. v [9].

V případě, že je k dispozici apriorní informace o funkci hustoty, je možné nalézt „optimální“ vyhlazovací parametr pomocí metody založené na *střední kvadratické chybě*²². Jak uvádí [10], str. 28, výraz pro MSE je možné dále upravovat, pomocí dalších kritérií jako např. *střední integrální kvadratická chyba*²³ a *asymptotická střední integrální kvadratická chyba*²⁴ a nakonec získat vzorec pro výpočet „optimálního“ parametru vyhlazení (pro odvození a důkaz viz [10], str. 33.)

$$h_{opt} = \left(\frac{R(k)}{(u_2(k))^2 R(f'')n} \right)^{\frac{1}{5}}, \quad (4)$$

kde $R(\bullet)$ označuje pro nějakou integrovatelnou funkci L funkcionál $R(L) = \int L^2(x) dx$,
 $u_2(\bullet)$ označuje pro nějakou funkci L funkcionál $u_2(L) = \int x^2 L(x) dx$ a
 f'' je druhá derivace známé funkce hustoty f .

Ve vzorci (4) figuruje skutečná funkce hustoty f . V praxi ale většinou žádná informace o funkci hustoty nebývá k dispozici. Bylo proto vyvinuto několik optimalizačních metod aproximujících h_{opt} , které využívají odhad funkce hustoty (blíže viz např. [10], [12] nebo [13]).

Necht' je náhodný výběr x_1, \dots, x_n nadále seřazen vzestupně $x_{(1)}, \dots, x_{(n)}$. „Vhodný“ vyhlazovací parametr je pak uvažován z předem definované množiny

$$H_n = \left\langle \min_{i=1, \dots, n-1} \{x_{(i+1)} - x_{(i)}\}; x_{(n)} - x_{(1)} \right\rangle, \quad (5)$$

kde $x_{(i)} \in \mathbb{R}$ je i -té pozorování ze seřazeného výběru náhodné veličiny X .

Podle [12], str. 29, patří mezi nejrozšířenější přístupy k odhadu h_{opt} metoda křížového ověřování nejmenších čtverců. Ta je založena na minimalizaci MISE. Vyhlazovací parametr pak může být volen pomocí vztahu ([12], str. 30)

$$\hat{h} = \arg \min_{h \in H_n} CV_f(h), \quad (6)$$

kde H_n je množina všech uvažovaných hodnot vyhlazovacího parametru podle (5) a
 $CV_f(\bullet)$ je funkce křížového ověřování pro funkci hustoty f definovaná vztahem (7).

Funkce křížového ověřování z (6) pro funkci hustoty má tvar (pro odvození viz [12], str. 29 nebo [14], str. 27)

$$CV_f(h) = \int_{-\infty}^{\infty} \hat{f}_h^2(t) dt - \frac{2}{n} \sum_{i=1}^n \hat{f}_{h,-i}(x_i) \quad (7)$$

²² [10], str. 27, definuje $MSE(\hat{f}_h(x)) = E(\hat{f}_h(x) - f(x))^2$

²³ [10], str. 28, definuje $MISE(\hat{f}_h(x)) = E\left(\int_{-\infty}^{\infty} (\hat{f}_h(x) - f(x))^2 dx\right)$.

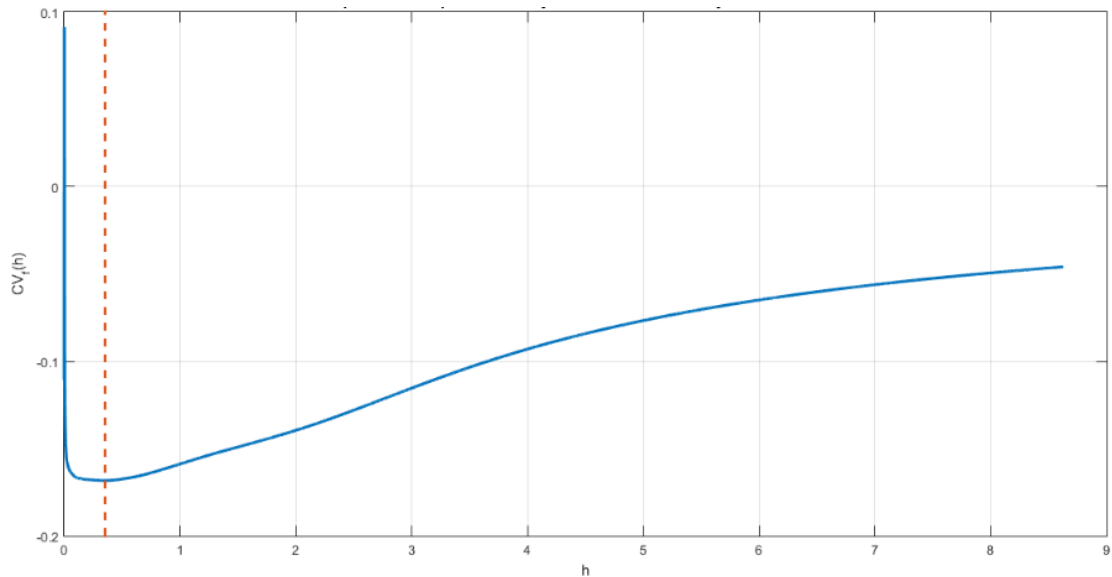
²⁴ Pro definici AMISE viz [9], str. 7.

kde $h \in \mathbb{R}^+$ je vyhlazovací parametr,

$\hat{f}_h(t)$ je jádrový odhad funkce hustoty v bodě t podle (3),

$\hat{f}_{h,-i}(x_i)$ je jádrový odhad funkce hustoty v bodě x_i podle (3) bez použití i -tého pozorování a $x_i \in \mathbb{R}$ je i -té pozorování náhodné veličiny X .

Obrázek 2.5 zachycuje průběh funkce křížového ověřování (7) s vyznačenou „vhodnou“ hodnotou vyhlazovacího parametru podle (6). h bylo „optimalizováno“ v rámci množiny $H_n \doteq (0, 8.6248)$. Minimalizace funkce křížového ověřování byla provedena pro stejná data jako v případě Obrázek 2.2.

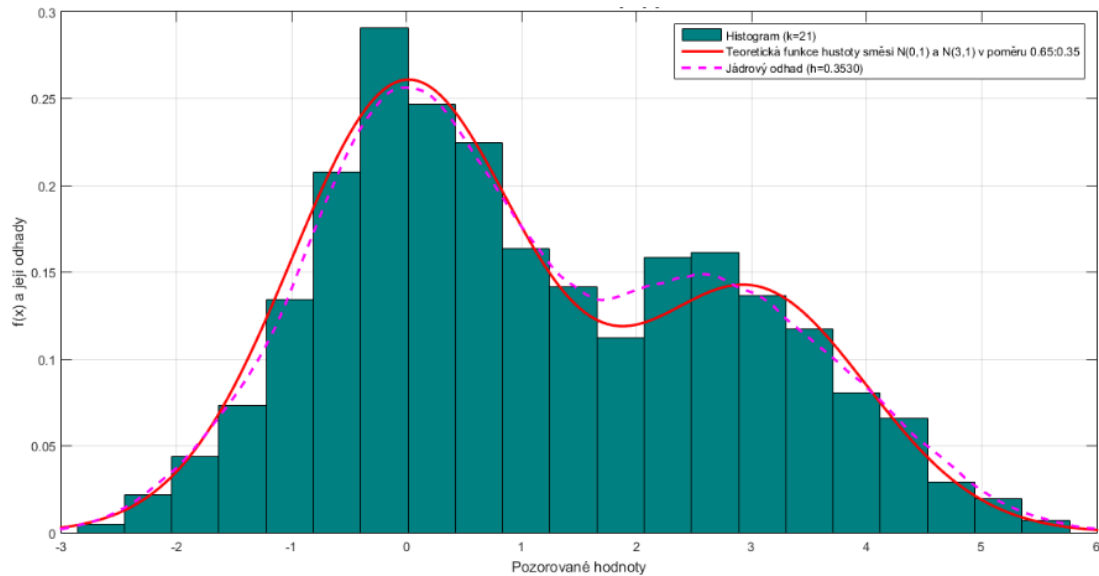


Obrázek 2.5: Průběh funkce křížového ověřování za použití Epanechnikovy jádrové funkce s vyznačenou minimální hodnotou $\hat{h} = 0.3530$, $n = 1000$, krok simulace nastaven na 0.0001 (zdroj: vlastní simulace)

2.1.3 Srovnání parametrického a neparametrického odhadu

Přibližné srovnání histogramu a jádrového odhadu hustoty zachycuje Obrázek 2.6. Pro stejná data jako v případě Obrázek 2.2 byl proveden odhad pomocí histogramu. Pro určení počtu tříd bylo použito Sturgesovo pravidlo²⁵. Pro porovnání obou přístupů byl také proveden jádrový odhad pomocí Epanechnikovy jádrové funkce z Tabulka 1. Parametr měřítka h byl zvolen pomocí metody křížového ověřování (6), viz Obrázek 2.5.

²⁵ se zaokrouhlením nahoru



Obrázek 2.6: Známa hustota a její odhady v podobě histogramu a jádrového odhadu za použití Epanechnikovy jádrové funkce a metody křížového odhadu vyhlazovacího parametru, $\hat{h} = 0.3530$, $n = 1\,000$ (zdroj: vlastní simulace)

2.2 Odhad neznámé distribuční funkce

Odhad distribuční funkce²⁶ lze provést sestavením empirické distribuční funkce, kdy každé pozorování má váhu $\frac{1}{n}$. Nevýhodou EDF je, že jde o schodovitou funkci (i v případě spojitého rozdělení náhodné veličiny) a nemusí tedy zachytit skutečný charakter distribuční funkce. Blíže k EDF viz [11], str. 13.

Možným přístupem k odstranění problému souvisejícího s EDF je využití jádrové distribuční funkce $K(x)$, kterou lze získat integrováním známé jádrové funkce hustoty $k(x)$ (např. z Tabulka 1) na jejím definičním oboru. Zde je patrná analogie ve vztahu teoretických funkcí, kdy distribuční funkce $F(x)$ se získá integrováním funkce hustoty $f(x)$ (viz [15], str. 63).

Na základě volby oblasti aplikace jádrových odhadů (viz kapitola 3) není jádrový odhad distribuční funkce předmětem této práce a jeho analýza nutná k pochopení této problematiky, a proto je v této podkapitole pouze zmíněn a blíže neanalyzován. Pro odvození vzorce pro jádrový odhad viz např. [2], str. 46. Přehled metod volby „vhodného“ vyhlazovacího parametru lze nalézt např. v [13], str. 33.

2.3 Odhad neznámé regresní křivky

Tato část vychází z [15] a [16].

Vzhledem ke splnění zadání práce je jejím hlavním cílem analýza problematiky regresních čar se zaměřením na použití neparametrického jádrového přístupu.

²⁶ Distribuční funkce $F(x)$ je funkce, pro kterou platí $F(x) = P(X \leq x)$. Tato funkce přiřazuje každému reálnému číslu pravděpodobnost, že náhodná veličina nabude hodnoty menší než toto číslo. (blíže viz [15], str. 57).

Obecně popisuje regresní křivka vztah mezi nenáhodnou²⁷ vysvětlující (nezávislou) proměnnou²⁸ X a náhodnou vysvětlovanou (závislou) proměnnou Y . Necht' je k dispozici n realizací (pozorování) proměnných X a Y v podobě uspořádaných dvojic $\{(x_i, y_i)\}_{i=1}^n$ a cílem²⁹ je aproximovat vztah mezi nimi pomocí nějaké křivky m . Tento vztah je popsán regresním modelem ve tvaru ([16], str. 3, nebo [17], str. 2)

$$y_i = m(x_i) + \varepsilon_i, \quad i = 1, \dots, n, \quad (8)$$

kde $y_i \in \mathbb{R}$ je i -té pozorování proměnné Y ,
 $x_i \in \mathbb{R}$ je i -té pozorování proměnné X ,
 $m(x)$ je hodnota neznámá regresní funkce v bodě x a
 $\varepsilon_i \in \mathbb{R}$ je i -té pozorování neznámé náhodné proměnné ε představující nesystematickou složku modelu (náhodnou chybu), na které jsou kladeny předpoklady (9).

Obecně se tento model zapisuje ve tvaru $Y = m(X) + \varepsilon$. [15], str. 363, uvádí předpoklady³⁰ pro základní model regresní analýzy (8) - hodnoty x_i jsou „pevné“ (tj. nenáhodné), a tudíž i funkční hodnoty $m(x_i)$ jsou „pevné“, ε_i jsou náhodné veličiny, pro které platí

$$\begin{aligned} E(\varepsilon_i) &= 0, & i &= 1, 2, \dots, n, \\ D(\varepsilon_i) = E(\varepsilon_i^2) &= \sigma^2, & i &= 1, 2, \dots, n, \\ C(\varepsilon_j, \varepsilon_k) = E(\varepsilon_j \varepsilon_k) &= 0, & j, k &= 1, 2, \dots, n, \end{aligned} \quad (9)$$

kde $E(\varepsilon_i)$ je střední hodnota³¹ náhodné veličiny ε_i ,
 $D(\varepsilon_i), \sigma^2 \in \mathbb{R}^+$ je rozptyl³² náhodné veličiny ε_i a
 $C(\varepsilon_j, \varepsilon_k)$ je kovariance³³ náhodných veličin ε_j a ε_k .

V (9) se tedy předpokládá, že pozorování y_i nejsou zatížena systematickými chybami (tj. nulová střední hodnota ε_i), pozorování y_i jsou naměřena se stejnou přesností (tj. shodný rozptyl σ^2 chybové složky modelu) a jednotlivé chyby měření jsou nekorelované (tj. lineárně navzájem nezávislé). Ze splnění podmínek (9) vyplývá³⁴, že pro náhodné veličiny y_i platí $E(y_i) = m(x_i), i = 1, 2, \dots, n$.

Pouhým pohledem na bodový graf nelze přesně určit typ funkčního vztahu mezi oběma veličinami z důvodu např. neschopnosti odhalení odlehklých pozorování³⁵ nebo vybočujících bodů³⁶. Jak již bylo naznačeno v úvodu této kapitoly, k odhadu regresní funkce můžeme přistupovat dvěma způsoby – parametricky a neparametricky.

²⁷ Obecně nemusí být nenáhodná – blíže viz [1], str. 57. Pro potřeby této práce je však uvažována nenáhodná vysvětlující proměnná.

²⁸ Obecně může jít o vektor vysvětlujících proměnných X_i , kde $i = 1, \dots, n; n \in \mathbb{N}$.

²⁹ Mimo jiné tímto úkolem se zabývá regresní analýza (blíže viz [15], str. 362 nebo [18], str. 187).

³⁰ Uváděny jsou pouze předpoklady pro jednorozměrný případ.

³¹ Střední hodnotu spojitě náhodné veličiny X s funkcí hustoty $f(x)$ definuje [15], str. 86:
 $E(X) = \int_{-\infty}^{\infty} xf(x) dx$.

³² Rozptyl náhodné veličiny X definuje [15], str. 91: $D(X) = E(X^2) - [E(X)]^2$.

³³ Kovarianci náhodných veličin (X, Y) definuje [15], str. 96: $C(X, Y) = E\{[X - E(X)][Y - E(Y)]\}$.

³⁴ Viz pravidla pro práci se středními hodnotami (např. [15], str. 86-87).

³⁵ Odlehlá pozorování (angl. „outliers“) definuje [1], str. 76, jako neočekávané hodnoty vysvětlované proměnné.

³⁶ Vybočující body (angl. „leverage points“) definuje [1], str. 76, jako neočekávané hodnoty vysvětlující proměnné.

V parametrickém případě se uvažuje konkrétní funkční předpis funkce m . Dále se předpokládá, že funkční vztah je plně popsán konečnou množinou parametrů (viz [16], str. 4).

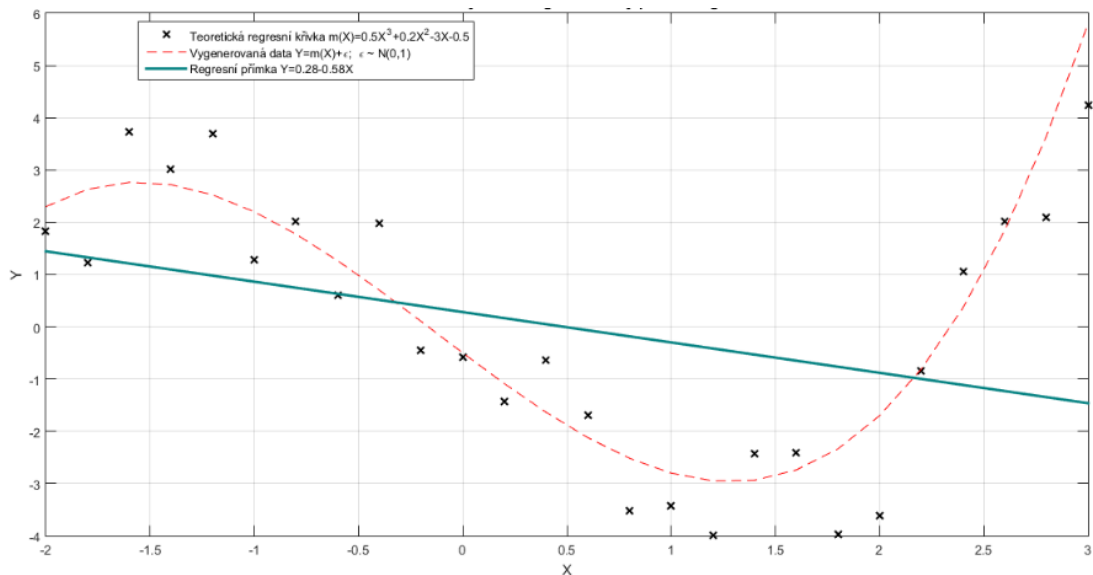
Neparametrické vyhlazování poskytuje flexibilní nástroje při analýze neznámého vztahu mezi vysvětlovanou a vysvětlující veličinou (viz [16], str. 4). Mezi výhody neparametrického přístupu řadí [1], str. 154, možnost provádět predikce nezávisle na nějakém určitém parametrickém modelu, vyhlazovat data, poměrně snadno se vyrovnat s chybějícími pozorováními, dále poskytuje kontrolu o výskytu odlehlých pozorování v datech, atd. Jak bylo naznačeno v úvodu této kapitoly, namísto parametrů např. ve smyslu regrese je v tomto případě odhadovaným parametrem přímo regresní funkce $m(X)$ z modelu (8).

2.3.1 Regrese

Tato část vychází z [1], [16], [18] a [19].

Metoda lineární³⁷ (příp. nelineární) regrese³⁸ (blíže viz např. [1], str. 38 a 57, [19], str. 81 nebo [18], str. 111) se používá pro odhad neznámých parametrů ve formě koeficientů předpokládané funkce m . Nejčastěji používanou metodou odhadu těchto parametrů je metoda nejmenších čtverců (blíže viz např. [15], str. 364).

Ne vždy se ale předpokládá správný funkční předpis, jelikož např. jsou ke zpracování zadána neznámá data, o kterých nejsou k dispozici žádné upřesňující informace, a automatickým použitím regrese by tudíž ani získaný odhad nebyl „vhodný“. Tuto situaci lze vidět na Obrázek 2.7. V tomto případě byl předpokládán vztah typu obecná přímka podle [18], str. 191, tj. model (8), kde $m(X) = \beta_0 + \beta_1 X$, kde β_0 (posunutí přímky po vertikální ose) a β_1 (směrnice přímky) jsou parametry modelu. Aproximace byla provedena pomocí zmiňované metody nejmenších čtverců.



Obrázek 2.7: Nevhodná volba regresního modelu – srovnání odhadu a teoretické regresní křivky, $n = 26$ (zdroj: vlastní simulace)

V situaci popsané výše je doporučován následující postup (podle [1], str. 153). Nejprve by měla být aplikována některá z neparametrických metod pro získání přibližné (počáteční) představy

³⁷ Myšleno lineární vzhledem k parametrům – např. $y = ax^2 + bx + c$.

³⁸ Regresi definuje [1], str. 57, jako aproximaci vztahu mezi vysvětlovanou a vysvětlující proměnnou.

o průběhu odhadované funkce a tedy o tom, jaký model by bylo „rozumné“ použít, než náhodně zkoušet různé modely, kdy navíc mohou nastat problémy s ověřením předpokladů použitých metod a s určením a ověřením vlastností získaných odhadů. Poté již může být podstatně snazší zvolit třídu parametrických modelů, které by data „vhodně“ aproximovaly, a z nich „optimální“ model.

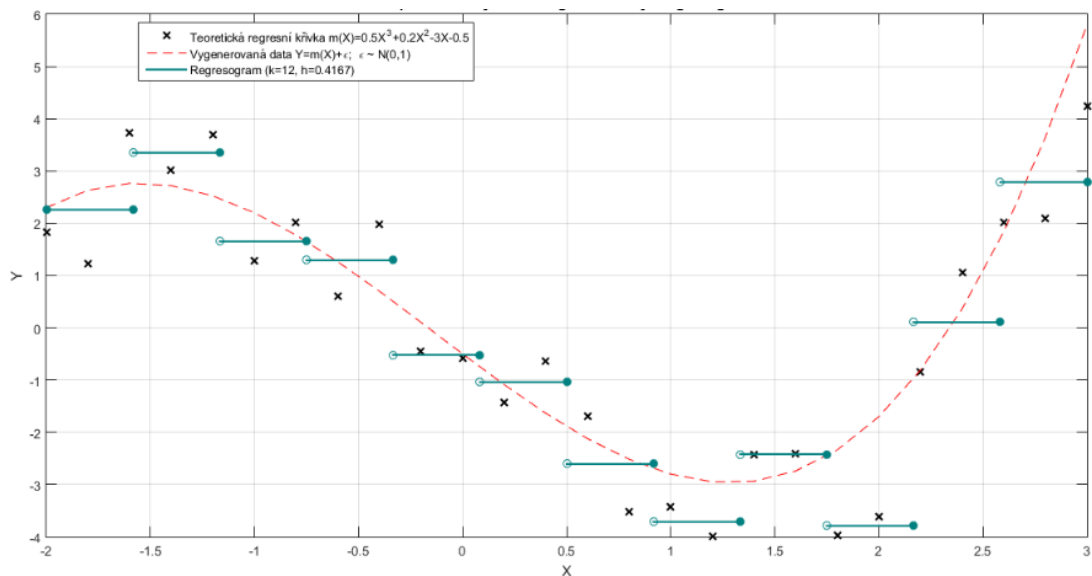
2.3.2 Regresogram

Tato část vychází z [1], [16] a [20].

Podle [1], str. 155, je nejjednodušší neparametrickou metodou odhadu regresní křivky zkonstruování regresogramu, který je přímou analogií histogramu pro odhad hustoty (či distribuční funkce (např. kumulovaný histogram)).

Pro vytvoření regresogramu lze použít následující postup. Horizontální osa je rozdělena na intervaly způsobem definovaným v tomto odstavci. Necht' $x_i, i = 1, \dots, n$ jsou pozorování ze spojitě náhodné veličiny X , dále existuje číslo $k \in \mathbb{N}$ a dělicí body $a_j \in \mathbb{R}$, kde $j = 0, 1, \dots, k$. Pak pro $a_0 = \min(x_i) < a_1 < \dots < a_{k-1} < a_k = \max(x_i)$ existují disjunktní³⁹ intervaly B_j , pro které platí $B_1 = (a_0; a_1)$ a $B_j = (a_{j-1}; a_j)$, $j = 2, 3, \dots, k$. Číslo k označuje volitelný počet intervalů.

Odhad v bodě $x \in B_j$ je proveden výpočtem aritmetického průměru těch hodnot vysvětlované proměnné Y , jejichž odpovídající hodnoty vysvětlující proměnné X leží ve stejném intervalu B_j . Všechny hodnoty v rámci intervalu jsou tedy aproximovány konstantou. Příklad regresogramu se stejně širokými intervaly, jejichž počet byl určen pomocí Sturgesova pravidla, uvádí Obrázek 2.8.



Obrázek 2.8: Odhad regresní křivky pomocí regresogramu s ekvidistantními intervaly, $h = 0.4167$, $n = 26$ (zdroj: vlastní simulace)

Podle [20], str. 15, by měl být volen „dostatečně velký“ počet intervalů pro minimální zkreslení (resp. rozptyl) odhadu ale ne „moc velký“, aby nedošlo ke ztrátě redukce rozptylu celkového odhadu, ke které dochází díky transformaci dat na intervalové průměry. Jelikož se jedná o přímou analogii histogramu, je pro určení počtu (resp. šířky⁴⁰) intervalů možné použít stejná pravidla jako

³⁹ Dvě množiny A_1 a A_2 jsou disjunktní, pokud platí: $A_1 \cap A_2 = \emptyset$, kde \emptyset je prázdná množina. [58].

⁴⁰ v případě stejně dlouhých intervalů

v případě histogramu (tj. určení počtu intervalů pomocí Sturgesova nebo Yuleova pravidla (viz str. 3), resp. šířky intervalů pomocí Scottova nebo Freedmanova a Diaconisova pravidla – viz str. 5).

V praxi však většinou, jak uvádí [1], str. 156, může být v každém intervalu různý počet pozorování. To vede k nepřesnosti odhadu. Z toho důvodu se používá tzv. regresogram s náhodným krokem, kdy počet intervalů se určí pomocí stejných pravidel jako v předchozím případě. Horizontální osa se rozdělí na k disjunktních intervalů tak, aby v každém byl přibližně stejný počet pozorování, tj. n/k .

Odhad regresní křivky $m(X)$ pomocí regresogramu má charakter skokové funkce, a tudíž se jedná pouze o hrubý odhad. Jak uvádí [16], str. 80, regresogram lze považovat za speciální případ jádrového odhadu $m(X)$ (při použití Parzenovy jádrové funkce – viz Tabulka 1) počítán pro střední body intervalů. Přestože se v podstatě jedná o jádrový odhad, vždy jde o nespojitou skokovou funkci, která nemusí „skutečný“ průběh $m(X)$ zachytit.

Podle [1], str. 155, se jedná o vysoce nerobustní⁴¹ odhad. Jeho nerobustnost je značně ovlivněna použitím výběrového aritmetického průměru. Dále mezi nevýhodami zmiňuje, že se jedná o vychýlený⁴² odhad se značným rozptylem. Přesto však je tato metoda velice užitečná díky výpočetní rychlosti a jednoduchosti a dostačující k získání prvotní představy o analyzovaných datech. Nerobustnost odhadu navíc může pozitivně přispět k analýze dat, jelikož může upozornit na možný výskyt odlehleho pozorování. Pro získání „hladšího“ odhadu je však třeba zvolit jiný přístup.

2.3.3 Klouzavé průměry

Tato část vychází z [17] a [21].

Metoda klouzavých průměrů přímo vychází z myšlenky regresogramu, jehož je zobecněním (viz [2], str. 12). V literatuře (např. [17], str. 23) je považována za nejtypičtější vyhlazovací techniku pro jednodimenzionální ekvidistantní data.

V této části je popsán jednoduchý⁴³ klouzavý průměr. Délka klouzavého okna (tj. délka intervalu, přes který vypočítáme průměr) je značena $h = 2p + 1$ a je určena volbou parametru p . Odhad v bodě x_i je průměr z lichého⁴⁴ počtu (h) pozorování, tj. z množiny $\{y_j: (i - p) \leq j \leq (i + p)\}$. Výsledný odhad má tvar (viz [17], str. 24)

$$\hat{m}_{SMA}(x_i) = \sum_{j=1-p}^{n+p} w_{ij} y_j, \quad (10)$$

kde $y_j \in \mathbb{R}$ je j -té pozorování náhodné proměnné Y odpovídající bodu x_j náhodné proměnné X ,
 $p \in \mathbb{N}$ je parametr určující délku klouzavého okna h a

$$w_{ij} = \begin{cases} \frac{1}{2p + 1}, & \text{pokud } -p \leq (i - j) \leq p, \\ 0, & \text{jinak.} \end{cases}$$

⁴¹ Nerobustní odhad je odhad, který není robustní (viz poznámka pod čarou č. 14).

⁴² Odhad T parametru $\theta \in \mathbb{R}$ se nazývá vychýlený, platí-li $ET = \theta + b(\theta)$, kde funkce b není identicky rovna nule. Funkce b se nazývá vychýlení odhadu T v bodě θ . (blíže viz [19], str. 101)

⁴³ Existují i další typy klouzavého průměru – např. centrovaný, dvojitý nebo vážený (viz např. [21], str. 95-102).

⁴⁴ Pro vyhlazování se používají i okna sudé délky, např. u centrovaných klouzavých průměrů (viz [22], str. 286).

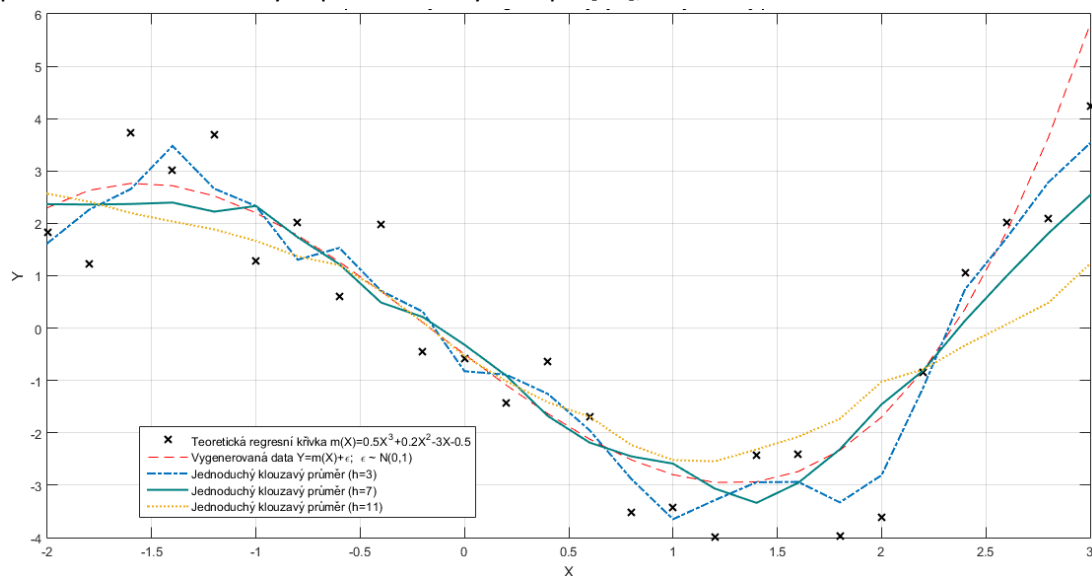
Z rovnice (10) je tedy zřejmé, že data jsou vyhlazována postupnými výpočty aritmetického průměru v jednotlivých oknech. Každé hodnotě v okně je tedy přiřazena stejná váha $\frac{1}{2p+1}$. Čím vyšší hodnota parametru p , tím „hladší“ je výsledný odhad křivky m (viz [17], str. 25) a tím větší je také pravděpodobnost odstranění náhodnosti v datech (viz [21], str. 94). To vede k volbě co nejdelšího okna. Na druhou stranu, čím delší okno je zvoleno, tím více pozorování (resp. informace) se v procesu průměrování „ztratí“, jelikož je potřeba $2p + 1$ pozorování k odhadu jedné hodnoty. Navíc volba delšího okna zapříčiní vyhlazení i malých, žádoucích výkyvů v datech.

Celé klouzavé okno se tedy nahrazuje hodnotou průměrnou a ta se přiřadí časovému okamžiku uprostřed tohoto okna. Tím ale zůstává neodhadnutých p počátečních a p koncových hodnot, tj. dohromady $2p$. Existuje několik způsobů, jak se s tímto problémem vypořádat. Jeden z nich uvádí [17], str. 24, kdy pro $j < 1$ nebo $j > n$ dodefinuje hodnoty y_j ve vzorci (10) tímto způsobem

$$\begin{aligned} y_0 &= y_1, & y_{-1} &= y_2, & \dots, & & y_{-p+1} &= y_p, \\ y_{n+1} &= y_n, & y_{n+2} &= y_{n-1}, & \dots, & & y_{n+p} &= y_{n-p+1}. \end{aligned} \quad (11)$$

V případě, že $p > n$, postup (11) se opakuje. Délka klouzavého průměru by podle [22], str. 281, měla být volena na základě sezónních nebo cyklických fluktuací, které je třeba vyhladit. Většinou se proto jedná o subjektivní posouzení charakteru dat, kdy si analytik sám určí přijatelnou „míru“ vyhlazení původních dat.

Pomocí klouzavého průměru se získá „hladší“ odhad regresní křivky než pomocí regresogramu. Příklad aproximace regresní křivky pomocí jednoduchého klouzavého průměru zachycuje Obrázek 2.9. Na obrázku lze porovnat odhady s různými délkami klouzavých oken. Detailněji se problematikou klouzavých průměrů zabývá např. [22], str. 274.



Obrázek 2.9: Odhad regresní křivky pomocí tří jednoduchých klouzavých průměrů různých délek, $h_1 = 3$, $h_2 = 7$, $h_3 = 11$, $n = 26$ (zdroj: vlastní simulace)

2.3.4 „Jádrový“ přístup

Tato část vychází z [2], [3], [16] a [20].

Jádrové odhady regresní křivky řadí [11], str. 71, mezi metody lokální neparametrické regrese.

Nadále bude předpokládáno, že $x \in \mathbb{R}$ je libovolná hodnota a bude uvažován regresní model ve tvaru (8). Metody tohoto typu vychází z myšlenky (viz [11], str. 71), že místo aritmetického průměru jako ve vzorci (10) použijeme vážené průměry z těch pozorování, které leží v symetrickém okolí bodu x (tj. $x \pm \frac{h}{2}$) a jádrová funkce přidělí největší váhu pozorováním ležícím „blízko“ x .

Existuje více různých variant použití jádrové funkce k odhadu regresní křivky. Nejznámější z nich uvádí [20], str. 158-160. Mezi ně patří odhady založené na práci Priestleyho a Chaa (1972), Nadaraya a Watsona (1964), Gassera a Müllera (1979) a Chenga a Lina (1981). Všechny čtyři uvedené odhady mají velmi podobné vlastnosti při ekvidistantních bodech pozorování $x_i, i = 1, \dots, n$. Např. v případě neekvidistantních bodů pozorování je ale „vhodnější“, jak [20] dále uvádí, použití Gasser-Müllerova odhadu než ostatních zmíněných odhadů. V uvedené literatuře je tato problematika zpracována podrobněji.

Tato práce je, z důvodu následného praktického použití, zaměřena na data s ekvidistantními body pozorování. Nejznámější (viz [14], str. 32) a ve zdrojích, ze kterých bylo čerpáno, nejčastěji (např. [1], str. 156, [3], str. 89, [11], str. 71, [16], str. 32, [23], str. 667, [22], str. 411 nebo [17], str. 133) popisován je Nadaraya-Watsonův odhad. Tento odhad je proto dále analyzován.

2.3.4.1 Nadaraya-Watsonův jádrový odhad regresní křivky

Tato část vychází z [24], str. 102.

Pro účely odvození tohoto odhadu je předpokládán i. i. d. výběr $\{(x_1, y_1), \dots, (x_n, y_n)\}$ ze spojitého rozdělení pravděpodobnosti s dvourozměrnou funkcí hustoty $f(x, y)$, kde $x, y \in \mathbb{R}$. Pozorování y_j jsou vysvětlované proměnné a x_j jsou vysvětlující proměnné. Pokud platí $E|y_j| < \infty$, pak existuje podmíněná střední hodnota

$$E(y_j|x_j) = g(x_j), \quad (12)$$

kde $g(\bullet)$ je Borelovsky měřitelná funkce na \mathbb{R} (blíže viz např. [25], str. 78).

Dále je zavedeno označení

$$e_j = y_j - g(x_j), \quad j = 1, 2, \dots, n, \quad (13)$$

kde $e_j \in \mathbb{R}$ je reziduum při odhadu y_j ,
 $y_j \in \mathbb{R}$ je j -té pozorování vysvětlované proměnné Y ,
 $x_j \in \mathbb{R}$ je j -té pozorování vysvětlující proměnné X a
 $g(\bullet)$ je Borelovsky měřitelná funkce na \mathbb{R} ze vztahu (12).

Tím se získá regresní model (8), tj. $y_j = g(x_j) + e_j$, kde z konstrukce tohoto tvaru platí podmínky (9). Funkce g tedy odpovídá regresní funkci m ze vztahu (8). Regresní model je určen funkcí g , která má tvar ([24], str. 102)

$$g(x) = \frac{\int_{-\infty}^{\infty} y f(x, y) dy}{f_X(x)}, \quad \text{pokud } f_X(x) > 0, \quad (14)$$

kde $f_X(x) = \int_{-\infty}^{\infty} f(x, y) dy$ je marginální⁴⁵ hustota $f(x, y)$, tj. funkce hustoty proměnné X .

⁴⁵ Blíže viz [15], str. 70.

Ze vztahu (14) je zřejmé, že odhad funkce $g(x)$ je možné získat odhadem funkcí hustoty f a f_X . Jak bylo popsáno v části 2.1.2.2, k odhadu neznámé jednorozměrné hustoty f_X lze použít jádrový odhad navržený Rosenblattem. Tento jádrový odhad s předem zvolenou jádrovou funkcí, např. z Tabulka 1, splňující podmínky (2) uvádí vztah (3). Nadále je podle [24], str. 103, předpokládáno, že pro parametr vyhlazení platí $\lim_{n \rightarrow +\infty} h = 0$ a $\lim_{n \rightarrow +\infty} n h = \infty$. Za platnosti podmínek (2) pro jádrovou funkci a limit pro parametr vyhlazení (uvedeny v předchozí větě) je jádrový odhad hustoty konzistentní⁴⁶ v každém bodě funkce $f_X(x)$ a tudíž platí

$$\sup_{x \in D(f_X)} f_X(x) < \infty, \quad (15)$$

kde $\sup_{x \in D(f_X)} f_X(x)$ označuje supremum funkce $f_X(x)$.

Důkaz tohoto tvrzení, podle [24], str. 103, využívá větu o postačující podmínce konzistence⁴⁷ odhadu. Stačí tedy ověřit splnění podmínek asymptotické nestrannosti odhadu⁴⁸, tj. $\lim_{n \rightarrow \infty} E(\hat{f}_X(x)) = f_X(x)$, a ověřit platnost vztahu $\lim_{n \rightarrow \infty} D(\hat{f}_X(x)) = 0$.

Důkaz. (Konzistence odhadu $\hat{f}_X(x)$). Na základě postačující podmínky konzistence odhadu $\hat{f}_X(x)$ je dokázána jeho asymptotická nestrannost, tedy

$$\begin{aligned} E(\hat{f}_X(x)) &= \int_{-\infty}^{\infty} f_X(t) \frac{k\left(\frac{x-t}{h}\right)}{h} dt \\ &= {}_a \int_{-\infty}^{\infty} f_X\left(x - \underbrace{h}_{\rightarrow 0} z\right) k(z) dz \\ &\xrightarrow{n \rightarrow \infty} f_X(x) \int_{-\infty}^{\infty} k(z) dz = f_X(x). \end{aligned} \quad (16)$$

V důkazu (16) se využívá předpoklad $\lim_{n \rightarrow +\infty} h = 0$ a druhá podmínka z (2). Symbolem $=_a$ je při výpočtu označeno místo, kde je použita substituce $z = \frac{x-t}{h}$. Pro splnění druhé části postačující podmínky je nutné ověřit konvergenci rozptylu odhadu $\hat{f}_X(x)$, tedy

⁴⁶ „Statistika g se nazývá konzistentním odhadem parametrické funkce $\gamma(\theta)$, jestliže pro libovolné $\varepsilon > 0$ platí $\lim_{n \rightarrow \infty} P(|g - \gamma(\theta)| < \varepsilon) = 1$ pro každé $\theta \in \Omega$, tj. jestliže pro $n \rightarrow \infty$ konverguje g v pravděpodobnosti k $\gamma(\theta)$.“ [15], str. 214.

⁴⁷ Postačující podmínka konzistence. Je-li g nestranný odhad nebo alespoň asymptoticky nestranný odhad parametrické funkce $\gamma(\theta)$ a rozptyl tohoto odhadu konverguje k nule, potom je odhad g konzistentní. (viz [15], str. 214, a [6], str. 163).

⁴⁸ „Statistika g se nazývá asymptoticky nestranným odhadem parametrické funkce $\gamma(\theta)$, jestliže $\lim_{n \rightarrow \infty} E(g) = \gamma(\theta)$ pro každé $\theta \in \Omega$.“ [15], str. 213.

$$\begin{aligned}
 D(\hat{f}_X(x)) &= D\left(\frac{1}{nh} \sum_{i=1}^n k\left(\frac{x-x_i}{h}\right)\right) = \frac{1}{n^2} \sum_{i=1}^n D\left(\frac{k\left(\frac{x-x_i}{h}\right)}{h}\right) \\
 &= \frac{1}{n^2} n D\left(\frac{k\left(\frac{x-x_i}{h}\right)}{h}\right) \\
 &= \frac{1}{n} E \frac{k^2\left(\frac{x-x_i}{h}\right)}{h^2} - \frac{1}{n} \left(E \frac{k\left(\frac{x-x_i}{h}\right)}{h}\right)^2 \\
 &= \frac{1}{nh} \int_{-\infty}^{\infty} f_X(x_i) \frac{k^2\left(\frac{x-x_i}{h}\right)}{h} dx_i - \frac{1}{n} \left(\int_{-\infty}^{\infty} f_X(x_i) \frac{k\left(\frac{x-x_i}{h}\right)}{h} dx_i\right)^2 \\
 &= \underset{\substack{\rightarrow 0 \\ n \rightarrow \infty}}{=} \frac{1}{nh} \int_{-\infty}^{\infty} f_X\left(x - \frac{h}{n} z\right) k^2(z) dz - \frac{1}{n} \left(\int_{-\infty}^{\infty} f_X\left(x - \frac{h}{n} z\right) k(z) dz\right)^2 \\
 &\rightarrow 0.
 \end{aligned} \tag{17}$$

Symbolem $\underset{\substack{\rightarrow 0 \\ n \rightarrow \infty}}{=}$ je při výpočtu označeno místo, kde je použita substituce $z = \frac{x-x_i}{h}$. V posledním kroku důkazu je využito také druhého limitního předpokladu uvedeného výše, tj. $\lim_{n \rightarrow +\infty} nh = \infty$. ■

Vztahy (16) a (17) dokazují, že jádrový odhad $\hat{f}_X(x)$ je konzistentním odhadem hustoty $f_X(x)$. Dále je k odhadu funkce g z (14) potřeba zkonstruovat odhad sdružené funkce hustoty $f(x, y)$, za podmínky, že $\hat{f}_X(x)$ je odhadem marginální hustoty $f(x, y)$ a výraz $\int_{-\infty}^{\infty} y \hat{f}(x, y) dy$ je nahrazen výrazem obsahujícím stejnou jádrovou funkci hustoty k jako v (3). Tento jádrový odhad hustoty $f(x, y)$ má tvar

$$\hat{f}(x, y) = \frac{1}{n} \sum_{i=1}^n \frac{k_*\left(\frac{x-x_i}{h}, \frac{y-y_i}{h}\right)}{h^2}, \tag{18}$$

kde $x_i \in \mathbb{R}$ je i -té pozorování vysvětlující proměnné X ,
 $y_i \in \mathbb{R}$ je i -té pozorování vysvětlované proměnné Y ,
 $k_*(x, y)$ je dvourozměrná jádrová funkce splňující podmínky (19) a
 $h \in \mathbb{R}^+$ je vyhlazovací parametr.

Vícerozměrnými jádrovými odhady se blíže zabývá např. [2], str. 55, nebo [8], str. 161. Pro potřeby odvození je v této práci pouze předpokládáno, že pro dvourozměrnou jádrovou funkci k_* v (18) platí

$$\int_{-\infty}^{\infty} y k_*(x, y) dy = 0 \quad \text{a} \quad \int_{-\infty}^{\infty} k_*(x, y) dy = k(x). \tag{19}$$

Potom $\hat{f}_X(x)$ je marginální hustota funkce $\hat{f}(x, y)$ a navíc platí

$$\int_{-\infty}^{\infty} y \hat{f}(x, y) dy = \frac{1}{n} \sum_{i=1}^n y_i \frac{k\left(\frac{x-x_i}{h}\right)}{h}. \tag{20}$$

Důkaz.

$$\begin{aligned}
 \int_{-\infty}^{\infty} y \hat{f}(x, y) dy &= \int_{-\infty}^{\infty} y \frac{1}{n} \sum_{i=1}^n k_* \left(\frac{x - x_i}{h}, \frac{y - y_i}{h} \right) dy \\
 &= \frac{1}{n} \int_{-\infty}^{\infty} \sum_{i=1}^n y k_* \left(\frac{x - x_i}{h}, \frac{y - y_i}{h} \right) dy \\
 &= \frac{1}{n} \sum_{i=1}^n \int_{-\infty}^{\infty} y k_* \left(\frac{x - x_i}{h}, \frac{y - y_i}{h} \right) dy \\
 &= {}_c \frac{1}{n} \sum_{i=1}^n \frac{1}{h} \int_{-\infty}^{\infty} (y_i + hz) k_* \left(\frac{x - x_i}{h}, z \right) dz \\
 &= \frac{1}{n} \sum_{i=1}^n \frac{y_i}{h} \int_{-\infty}^{\infty} k_* \left(\frac{x - x_i}{h}, z \right) dz + \frac{h}{h} \int_{-\infty}^{\infty} z k_* \left(\frac{x - x_i}{h}, z \right) dz \\
 &= \frac{1}{n} \sum_{i=1}^n y_i \frac{k \left(\frac{x - x_i}{h} \right)}{h}
 \end{aligned} \tag{21}$$

Symbolem $=_c$ je při výpočtu označeno místo, kde je použita substituce $z = \frac{y - y_i}{h}$. V posledním kroku důkazu je využito předpokladů (19). ■

Tudíž dosazením (3) a (20) do vztahu (14) se získá odpovídající odhad regresní funkce

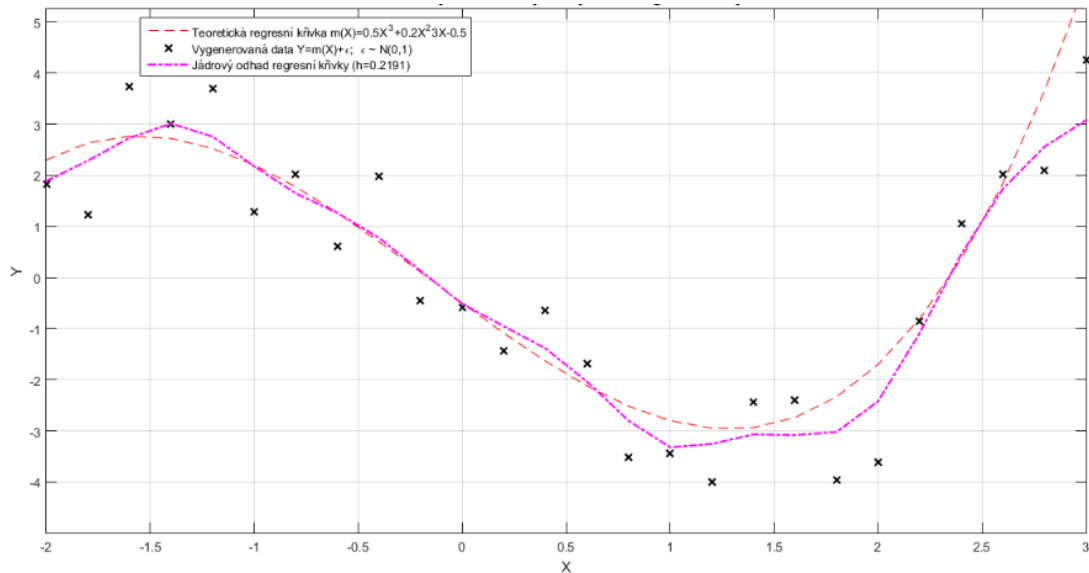
$$\hat{g}(x) = \hat{m}_h(x) = \begin{cases} \frac{\frac{1}{n} \sum_{i=1}^n k \left(\frac{x - x_i}{h} \right) y_i}{\frac{1}{n} \sum_{j=1}^n k \left(\frac{x - x_j}{h} \right)}, & \sum_{j=1}^n k \left(\frac{x - x_j}{h} \right) \neq 0, \\ 0, & \text{jinak,} \end{cases} \tag{22}$$

kde $\hat{g}(x)$ je odhad regresní křivky g z modelu vycházejícího ze vztahu (13),
 $\hat{m}_h(x)$ je jádrový odhad regresní křivky m z modelu (8) při zvoleném parametru vyhlazení h ,
 $h \in \mathbb{R}^+$ je parametr vyhlazení (viz [20], str. 157),
 $k(x)$ je jádrová funkce splňující podmínky (2),
 $x_i \in \mathbb{R}$ je i -té pozorování vysvětlující proměnné X a
 $y_i \in \mathbb{R}$ je i -té pozorování vysvětlované proměnné Y .

Vztah (22) se nazývá Nadaraya-Watsonův jádrový odhad regresní křivky. Tento odhad regresní křivky je vážený průměr pozorování závisle proměnných y_i . Vztah (22) také zajišťuje, že součet těchto vah je roven jedné (viz⁴⁹ [20], str. 160). Váhy jsou přiřazeny jednotlivým pozorováním y_i na základě vybrané jádrové funkce. Např. při volbě některé jádrové funkce z Tabulka 1 a odhadu regresní křivky v bodě x platí, že čím blíže je x k bodu x_i , tím větší váha je přiřazena pozorování y_i .

Příklad použití Nadaraya-Watsonova odhadu za použití Epanechnikovy jádrové funkce z Tabulka 1 a níže popsané metody křížového odhadu (29) „vhodného“ vyhlazovacího parametru ukazuje Obrázek 2.10.

⁴⁹ v publikaci rovnice pod označením (4.9)



Obrázek 2.10: Nadaraya-Watsonův jádrový odhad regresní křivky za použití Epanechnikovy jádrové funkce a metody křížového odhadu vyhlazovacího parametru, $\hat{h} = 0.2191$, $n = 26$ (zdroj: vlastní simulace)

2.3.4.2 Rozbor vlivu parametru vyhlazení na výsledný odhad

Tato část vychází z [3], [16] a [20].

Pro analýzu volby parametru vyhlazení lze vztah (22) přepsat do tvaru (vycházíme z [3], str. 89)

$$\hat{m}_h(x) = \frac{1}{n} \sum_{i=1}^n \left(\frac{k\left(\frac{x-x_i}{h}\right)}{\frac{1}{n} \sum_{j=1}^n k\left(\frac{x-x_j}{h}\right)} \right) y_i = \frac{1}{n} \sum_{i=1}^n w_i(x, h) y_i, \quad (23)$$

kde $h \in \mathbb{R}^+$ je vyhlazovací parametr,

$k(x)$ je jádrová funkce splňující podmínky (2),

$x_i \in \mathbb{R}$ je i -té pozorování vysvětlující proměnné X ,

$y_i \in \mathbb{R}$ je i -té pozorování vysvětlované proměnné Y a

$$w_i(x, h) = \frac{k\left(\frac{x-x_i}{h}\right)}{\frac{1}{n} \sum_{j=1}^n k\left(\frac{x-x_j}{h}\right)} \quad \text{označuje váhovou funkci přiřazující váhy } y_i.$$

Vyhlazovací parametr, jak poznamenává [3], str. 90, určuje stejně jako v případě jádrového odhadu hustoty (viz část 2.1.2 v této práci) stupeň „vyhlazení“ odhadu $\hat{m}_h(x)$, jelikož podle [20], str. 158, řídí špičatost váhové (tj. jádrové) funkce a tím „stupeň“ závislosti na pozorováních „blízko“ x .

Při následujícím odvozování je vycházeno ze vztahu (23). V případě prvního extrému, tj. $h \rightarrow 0$, a za předpokladu, že odhad je prováděn pouze pro $x = x_i$, $i = 1, \dots, n$, tj. jinde není definován, platí $w_i(x, h) \xrightarrow{h \rightarrow 0} n$.

Důkaz. Pro pevné h a pro $i = 1, \dots, n$, kde $n \in \mathbb{N}$ je počet pozorování, platí

$$w_i(x_i, h) = \frac{k\left(\frac{x_i-x_i}{h}\right)}{\frac{1}{n} \sum_{j=1}^n k\left(\frac{x_i-x_j}{h}\right)} \xrightarrow{h \rightarrow 0} \frac{k(0)}{\frac{1}{n} \left[\sum_{j=1}^{i-1} 0 + k(0) + \sum_{j=i+1}^n 0 \right]} = \frac{k(0)}{\frac{1}{n} k(0)} = n. \quad (24)$$

Na základě důkazu (24) proto odhad $\hat{m}_h(x_i)$ v bodě x_i konverguje k hodnotě y_i , tj. (viz [16], str. 33)

$$\hat{m}_h(x_i) \xrightarrow{h \rightarrow 0} \frac{1}{n} y_i = y_i \quad (25)$$

Výsledný odhad tedy v podstatě může kopírovat data a provádíme interpolaci těchto dat. V případě druhého extrému, kdy $h \rightarrow \infty$, platí $w_i(x, h) \xrightarrow{h \rightarrow \infty} 1$ pro všechna x .

Důkaz. $\forall x \in \mathbb{R}$, pro pevné h a pro $i = 1, \dots, n$, kde $n \in \mathbb{N}$ je počet pozorování, platí

$$w_i(x, h) = \frac{\overbrace{k\left(\frac{x-x_i}{h}\right)}^{\rightarrow 0}}{\underbrace{\frac{1}{n} \sum_{j=1}^n k\left(\frac{x-x_j}{h}\right)}_{\rightarrow 0}} \xrightarrow{h \rightarrow \infty} \frac{k(0)}{\frac{1}{n} \left[\sum_{j=1}^n k(0) \right]} = \frac{k(0)}{\frac{1}{n} n k(0)} = 1. \quad (26)$$

■

Na základě důkazu (26) proto odhad $\hat{m}_h(x)$ konverguje k výběrovému průměru ze všech pozorování veličiny Y , tj. (viz [16], str. 33)

$$\hat{m}_h(x) \xrightarrow{h \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n y_i. \quad (27)$$

Získaný odhad je tedy konstantní funkce pro všechna x .

Stěžejní úlohou je tedy opět volba parametru h tak, aby byl „vhodným“ kompromisem mezi extrémně podhlazeným odhadem (25) a extrémně přehlazeným odhadem (27). Jeho volba může být závislá na apriorní znalosti rozdělení.

Zjednodušeně to tedy znamená, že čím nižší parametr vyhlazení h je volen, tím více ovlivňují jádrový odhad pozorování blízká odhadované hodnotě a odhad křivky se spíše blíží původním hodnotám y_i . Naopak čím vyšší je hodnota h , tím méně odhad zohledňuje jednotlivá pozorování, jelikož dochází k průměrování více hodnot, a tedy širšímu rozložení vah. [20], str. 158, zmiňuje, že parametr vyhlazení mj. určuje, jak vzdálená pozorování ještě budou mít vliv na odhad regresní křivky.

2.3.4.3 Volba vyhlazovacího parametru pro jádrový odhad regresní křivky

Tato část vychází z [4] a [16].

Stejně jako v případě optimalizace vyhlazovacího parametru pro jádrový odhad funkce hustoty v části 2.1.2.3, tak i pro jádrový odhad regresní křivky existuje několik metod volby „vhodného“ h . Ty lze rozdělit do dvou skupin podle toho, jestli je teoretická regresní křivka známá nebo ne. Optimalizace je uvažována v bodech měření $x_i, i = 1, \dots, n$.

V případě, že je k dispozici apriorní informace o regresní křivce, odvozuje [4], str. 25, vztah (28) pomocí globální chyby MSE, tj. *průměrná střední kvadratická chyba*⁵⁰. Teoretickou „optimální“ hodnotu vyhlazovacího parametru definuje vztahem

⁵⁰ [4], str. 24, definuje $AMSE(h) = \frac{1}{n} \sum_{i=1}^n E(\hat{m}_h(x_i) - m(x_i))^2$.

$$h_{opt} = \left(\frac{\sigma^2 V(k) (\kappa!)^2}{2\kappa n \beta_\kappa^2 A_\kappa} \right)^{\frac{1}{2\kappa+1}}, \quad (28)$$

kde $\sigma^2 \in \mathbb{R}^+$ je rozptyl náhodné složky regresního modelu (viz předpoklady (9)),
 $k(x)$ je jádrová funkce splňující podmínky (2),
 $\kappa \in \mathbb{N}_0$ je řád jádrové funkce (viz poznámka pod čarou 16),
 $m^{(\kappa)}(x)$ je κ -tá derivace teoretické regresní křivky m ,

$$V(k) = \int_{-1}^1 k^2(x) dx,$$

$$\beta_\kappa = \int_{-1}^1 x^\kappa k(x) dx \text{ a}$$

$$A_\kappa = \int_0^1 \left(m^{(\kappa)}(x) \right)^2 dx.$$

Parametry σ^2 a $m(x)$ ale většinou nebývají k dispozici, a proto bylo vyvinuto několik metod aproximujících h_{opt} ve (28) (blíže viz např. [4], [11] nebo [16]).

Podle [4], str. 28, patří i v tomto případě mezi nejznámější aproximační metody metoda křížového ověřování. Tento postup využívá odhady provedené bez použití postupně každého z pozorování a je založen na *reziduálním součtu čtverců*⁵¹.

Je předpokládáno, že „vhodný“ vyhlazovací parametr se nachází v předem definované množině H_n podle (5). Jak bylo uvedeno v úvodní části podkapitoly 2.3, jednotlivá měření y_i jsou získána v časových okamžicích x_i . Ty jsou již ze své podstaty seřazeny a není proto nutné (z důvodu splnění předpokladu pro (5)) uvažovat toto řazení a vztah je platný.

Vyhlazovací parametr h může být volen pomocí vztahu ([4], str. 28)

$$\hat{h} = \arg \min_{h \in H_n} CV_m(h), \quad (29)$$

kde H_n je množina všech uvažovaných hodnot vyhlazovacího parametru podle (5) a $CV_m(\bullet)$ je funkce křížového ověřování pro regresní funkci m definovaná vztahem (30).

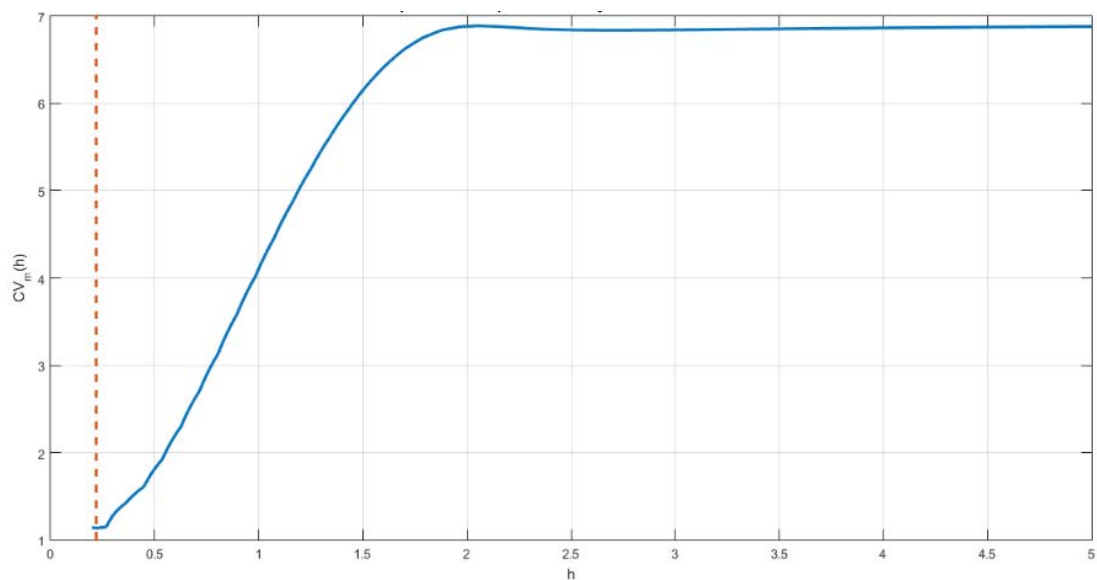
Funkce křížového ověřování z (29) pro regresní funkci má tvar ([4], str. 28)

$$CV_m(h) = \frac{1}{n} \sum_{i=1}^n [\hat{m}_{h,-i}(x_i) - y_i]^2, \quad (30)$$

kde $h \in \mathbb{R}^+$ je vyhlazovací parametr,
 $\hat{m}_{h,-i}(x_i)$ je Nadaraya-Watsonův jádrový odhad regresní funkce v bodě x_i podle (22) bez použití i -tého pozorování,
 $x_i \in \mathbb{R}$ je i -té pozorování vysvětlující proměnné X a
 $y_i \in \mathbb{R}$ je i -té pozorování vysvětlované proměnné Y .

Obrázek 2.11 zachycuje průběh funkce křížového ověřování (30) s vyznačenou „vhodnou“ hodnotou vyhlazovacího parametru podle (29). Jak je také vidět na obrázku, h bylo „optimalizováno“ v rámci množiny $H_n = \langle 0.2, 5 \rangle$. Minimalizace funkce křížového ověřování byla provedena pro stejná data jako v případě Obrázek 2.7.

⁵¹ [4], str. 25, definuje $RSS(h) = \frac{1}{n} \sum_{i=1}^n (\hat{m}_h(x_i) - y_i)^2$.



Obrázek 2.11: Průběh funkce křížového ověřování za použití Epanechnikovy jádrové funkce s vyznačenou minimální hodnotou $\hat{h} = 0.2191$, $n = 26$, krok simulace nastaven na 0.0001 (zdroj: vlastní simulace)

3 Obchodování s akcemi na BCPP

Tato část vychází především z [26], dále pak z [27], [28] a [29].

Poznatky především z části 2.3.4 budou aplikovány na data z burzy cenných papírů, jelikož se jedná o prakticky využitelnou oblast, kde by vytvořené modely, v případě dosažení pozitivních výsledků v této práci, mohly vést k jejich využití při vlastním obchodování. Zvoleny byly různorodé akciové tituly obchodované na *Burze cenných papírů Praha, a.s.*⁵² (pro princip výběru a popis titulů viz kapitola 5).

BCPP je spolu s burzami v Budapešti, Lublani a Vídni součástí největší burzovní skupiny ve střední a východní Evropě. BCPP se skládá ze tří burzovních trhů⁵³ – *Prime* (největší a nejprestižnější emise⁵⁴ akcií; nejpřísnější podmínky a pravidla pro zařazení emisí), *Standard* (velké a prestižní emise akcií, zároveň umožňuje přijetí emisí bez souhlasu emitenta⁵⁵, pokud jsou již obchodovány na jiném regulovaném trhu EU) a *Start* (pro malé a střední firmy). Každá emise akcií je zařazena do jednoho z těchto trhů. Oficiálním indexem⁵⁶ BCPP je *PX*, který se skládá z nejobchodovanějších blue chips⁵⁷ akcií. (blíže viz [30]).

Pod pojmem obchod (resp. uzavření obchodu) je chápáno vstoupení nebo vystoupení z dlouhé nebo krátké pozice⁵⁸ podle [27], str. 8-9. Vstup do dlouhé pozice znamená, že jsou nakoupeny cenné papíry za předpokladu, že jejich cena bude růst. Dlouhá pozice se otevírá nákupem cenných papírů a uzavírá jejich prodejem. Do krátké pozice se vstupuje „vypůjčením“ a prodejem cenných papírů od obchodníka za předpokladu, že jejich cena bude klesat. Očekává se, že později je bude možné nazpět levněji nakoupit a tím i navrátit původnímu majiteli.

Nezávisle na jakém trhu se obchoduje, platí určitá pravidla pro každého investora⁵⁹. Pro potřebu této práce je nutné stanovit určité předpoklady. V této práci je uvažován investor, který chce do akcií investovat větší obnos (např. 500 000 Kč) a co nejvýhodněji (ve smyslu maximalizace zisku) zhodnotit své finance.

Žádný investor nemůže s akcemi na burze obchodovat sám, jak uvádí [31], ale pouze prostřednictvím některého z členů burzy, v tomto případě členů BCPP⁶⁰. Z tohoto seznamu byla zvolena společnost *Fio banka, a.s.*⁶¹, která je podle článku [32] nejvýhodnějším obchodníkem při větších objemech obchodů.

⁵² <https://www.pse.cz/>

⁵³ Blíže viz <https://www.pse.cz/dokument.aspx?k=Schema-Trhu>.

⁵⁴ „Soubor vzájemně zastupitelných cenných papírů vydaných emitentem na základě jeho rozhodnutí. Emitent stanoví druh cenného papíru, nominální hodnotu, počet kusů, formu a podobu.“ [59]

⁵⁵ „Společnost, která vydá cenné papíry, především za účelem získání finančních prostředků.“ [60]

⁵⁶ „Ukazatel znázorňující trend vývoje na určitém akciovém trhu. Z tržních cen vybraných akcií se vytváří agregátní ukazatel, který odráží současný stav a vývoj kurzu na burze.“ [59]

⁵⁷ „Označení pro nejkvalitnější a nejvíce obchodované emise akcií na daném veřejném trhu.“ [60]

⁵⁸ „Postavení investora vzhledem k trhu daného cenného papíru.“ [27], str. 8.

⁵⁹ „Každá fyzická nebo právnická osoba, která investuje své volné peněžní prostředky, a to přímo nebo prostřednictvím investičního zprostředkovatele, do investičních nástrojů dané země.“ [59]

⁶⁰ Pro jejich kompletní seznam viz [31].

⁶¹ <http://www.fio.cz/>

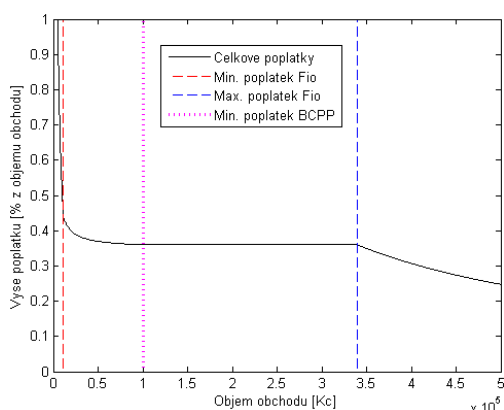
3.1 Poplatky spojené s obchodováním na BCPP

S každým uzavřeným obchodem, obecně na jakékoliv burze, se pojí poplatky, které jsou rozděleny na poplatky odvádějící se obchodníkovi (někdy také broker), tj. členovi burzy, u kterého byl založen obchodní účet, a trhu, tj. burze. V případě Fio je uvažováno obchodování výhradně přes aplikaci *e-Broker*⁶², skrze kterou je podání jakéhokoliv pokynu do systému zdarma (viz [28]). Poplatky jsou hrazeny až v případě realizace pokynu na trhu. Jednotlivé sazby poplatků při obchodování na českém burzovním trhu prostřednictvím Fio jsou shrnuty v Tabulka 2.

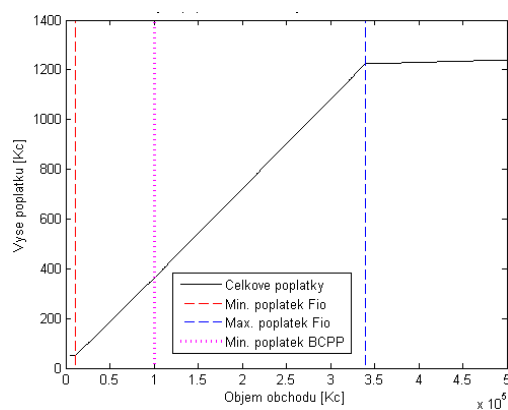
Poplatek při nákupu i prodeji	Procentní sazba (z objemu obchodu)	Omezení výše poplatku	
		Minimum	Maximum
Fio	0.35 %	40 Kč	1190 Kč
BCPP	0.01 %	10 Kč	4000 Kč

Tabulka 2: Přehled poplatků za uzavřené obchody s českými akciovými tituly při obchodování na BCPP prostřednictvím Fio (zdroje: [26], str. 5, [28] a [29])

Pro lepší představu o výši poplatků z Tabulka 2 ve vztahu k objemu obchodu, byly vytvořeny grafy zachycující tuto závislost, které jsou na Obrázek 3.1 a na Obrázek 3.2. Obrázek 3.1 zachycuje relativní výši poplatku v procentech z objemu obchodu a, pro porovnání, Obrázek 3.2 absolutní výši poplatku, tj. v českých korunách. Objem obchodu je uvažován v českých korunách. Kromě křivky představující vývoj výše poplatku jsou v grafech uvedeny také hraniční⁶³ objemy obchodů spočítané dle Tabulka 2.



Obrázek 3.1: Vývoj relativní výše poplatků při obchodování českých akciových titulů na BCPP prostřednictvím Fio (zdroj: vlastní simulace)



Obrázek 3.2: Vývoj absolutní výše poplatků při obchodování českých akciových titulů na BCPP prostřednictvím Fio (zdroj: vlastní simulace)

Na základě uvedených grafů lze říci, že pro minimalizaci poplatků při obchodování s akciemi českých akciových společností na BCPP prostřednictvím Fio je třeba maximalizovat objem obchodu.

3.2 Daň z příjmu

V případě výnosu z uzavřených obchodů s akciemi, je ze zákona povinné (za splnění určitých podmínek) odvádět státu 15% daň z příjmu.

Tato práce se zaměřuje na investice do akcií a obchodování s nimi, a zároveň se soustředí na kapitálový výnos z obchodů, tj. rozdíl mezi prodejní a pořizovací cenou. Bude počítáno s kurzy očištěnými o dividendu, a proto není zisk v podobě dividendy uvažován.

⁶² Pro detailnější informace viz <http://www.fio.cz/akcie-investice/e-broker>

⁶³ Horní mez poplatku burze není v grafech zanesena z důvodu její výše (je fixní až od objemu 40 mil. Kč).

Podle článku o akcích a daních [33] a také §4 první části (Daň z příjmů fyzických osob) Zákona č. 586/1992 Sb. o daních z příjmů [34] je povinností odvést daň z příjmu z realizovaného zisku, tedy kapitálového zisku očištěného o poplatky související s obchodem, v případě, že zisk z převodu cenného papíru přesáhne 100 000 Kč za rok nebo vlivem doby mezi nákupem a prodejem cenného papíru dojde k nesplnění tzv. časového testu.

Časový test je hraniční období, po jehož uplynutí je zisk od daně osvobozen. V roce 2014 došlo k prodloužení časového testu z 6 měsíců na 3 roky. Tato práce se však bude zaměřovat na krátkodobější obchodování (a zároveň bude investován veškerý dostupný kapitál na obchodním účtu⁶⁴ – dojde navíc i ke splnění první podmínky o výši zisku z převodu cenného papíru za rok), a proto budou veškeré zisky z obchodování uváděny hrubé, tj. před zdaněním.

⁶⁴ Blíže je použita strategie obchodování popsána v části 6.3.1.

4 Technická analýza akcií

Každý investor chce co „nejlépe“ (ve smyslu maximalizace zisku) zhodnotit své finance. Aby se tak stalo, musí se pokusit předpovědět, kdy je nejlepší okamžik investovat. Zároveň ale dobrý investor musí mít cit pro určení „vhodného“ času, kdy by jeho investice vedla k co nejvyššímu zisku a nestala se ztrátovou. V případě obchodování na akciovém trhu si může pomoci různými postupy a nástroji, jak takové okamžiky odhadnout. Jedním z nich je disciplína zvaná technická analýza, jejíž nástroje lze použít nejen na akciovém ale i na měnovém a komoditním⁶⁵ trhu. V každém případě je však nutné dbát na specifika obchodování na konkrétním trhu.

4.1 Definice technické analýzy

Tato část vychází z [27], [35], [36] a [37].

Průkopníkem technické analýzy byl *Charles Dow*, který jí položil základy. Sám pouze shrnoval své poznatky ze studování vývoje trhu. Jeho následovník *William Hamilton* rozšířil jeho technický přístup a až *Charles Rhea* dal ve 30. letech 20. století⁶⁶ pevnou formu Dowovým konceptům a shrnul je v knize pod názvem *Dow Theory*. (blíže viz [36] nebo [37], str. 233).

Technickou analýzu lze definovat jako metodu odhadu cen akcií nebo také studium, jak jednotlivých akciových titulů, tak i celkového trhu, které je založené na poměru nabídky a poptávky. Analytici, kteří ji používají pro odhady budoucího vývoje, většinou vycházejí při svých odhadech z trendů⁶⁷ a formací v grafech historických kurzů akcií, příp. do svých analýz zahrnují také objemy obchodů daných akciových titulů. Její nástroje napomáhají k identifikaci správných okamžiků pro obchodování, a lze ji proto považovat za dobrý nástroj a rádce pro situaci nastíněnou v úvodu této kapitoly.

Tato disciplína se opírá o následující tři základní principy (viz [27], str. 12 a [35], str. 13):

- 1) Tržní ceny odrážejí a zahrnují veškeré informace, které je mohou ovlivnit.
- 2) Kurzy se pohybují v trendech a trendy mají jistou míru setrvačnosti. (31)
- 3) Dění na trhu (vývojové cykly a formace) se opakují.

První předpoklad v (31) tedy uvádí, že tržní cena akcií je stanovena vztahem nabídky a poptávky a nemá smysl zabývat se fundamentálními faktory jako např. finanční výkazy firem a porovnávání tržní ceny s vnitřní hodnotou akcií. Tyto a zároveň i politické, psychologické a jakékoliv další faktory jsou již v ceně zahrnuty. Typický je nezájem o důvody, proč se něco děje, ale raději se zabývá tím, co se děje a jak na to reagovat.

Analýza vývoje tržních kurzů je podle druhého předpokladu v (31) to jediné, co je zapotřebí. Kurzy mají vždy buď rostoucí, klesající nebo konstantní trend a tento trend následují až do té doby, než dojde k jeho oslabení a poté i ke změně.

⁶⁵ „Trhy, na kterých se obchoduje s kontrakty na dodávku různých komodit – vzácné kovy, ropa, pšenice, kukuřice. Do finančních trhů se zařazují pouze v případě trhu s cennými kovy – zlato, stříbro a platina.“ ([60])

⁶⁶ 30 let po smrti C. Dowa

⁶⁷ „V technické analýze se jedná o tendenci vývoje ceny aktiva pohybovat se určitým směrem v čase. K jeho identifikaci slouží nástroje (grafické formace, indikátory) technické analýzy.“ ([60])

Třetí předpoklad v (31) říká, že, jelikož lidé mají tendenci reagovat v podobných situacích podobně jako v minulosti, je možné odhadovat budoucí vývoj na trhu na základě vývoje v minulosti.

Podle [27], str. 13, funguje tato metodika na principu dvou druhů informací – fundamentální informace a informace o chování trhu. Fundamentální informace jsou „*veškeré informace o firmě a jejím hospodaření, o makroekonomické situaci v zemi, o politice, o zahraničním obchodě a kursech, apod.*“ Informace o chování na trhu zahrnují „*veškeré informace o chování investorů, o jejich reakcích v určitých situacích v minulosti, o trendech, apod.*“ Technická analýza je pak založená na tom, že se všichni účastníci trhu snaží co „nejlépe“ zahrnout fundamentální informace do ceny, a tím vytvářejí informace o chování trhu. Tím, že investoři postupně zahrnují fundamentální informace do ceny, dochází na trhu ke stejným nebo podobným situacím, na které pak investoři stejně nebo podobně reagují.

V podstatě ji lze rozdělit do dvou oblastí podle toho, na základě čeho investor předpovídá budoucí vývoj kurzu:

- 1) studování grafických formací (grafické modely) a
- 2) analýza technických indikátorů (matematické modely).

4.2 Studování grafů

Na osu x jsou vynášeny obchodní dny⁶⁸ a na osu y odpovídající hodnoty. V grafu se hledají určité vzorce chování trhu (nebo také grafické formace), podle kterých získá investor představu o jeho budoucím vývoji díky předpokladu 3 v (31). Každá formace má svoji vlastní interpretaci. Tento přístup však není předmětem této práce, a proto se jím zde nebudeme blíže zabývat – pro detailnější popis viz např. [27], str. 15-49 nebo [35], str. 15-50.

4.3 Technické indikátory

Tato část vychází z [27].

Technické indikátory se využívají při definování obchodních systémů, které v reakci na tyto ukazatele generují signály k nákupu, k prodeji nebo k vyčkávání, tj. nenakupovat ani neprodávat. Obecně platí, že k identifikaci „optimálního“ okamžiku realizace obchodu je třeba zkombinovat více indikátorů.

V [27], str. 50, je indikátor definován jako „*funkce (resp. vektor funkcí), která pro každý obchodní den t , na němž je definována, přiřadí reálné číslo (resp. vektor reálných čísel), konstruované na základě znalosti cen nebo objemů konkrétní akcie do obchodního dne t a pro příslušný vektor parametrů π .*“ Technické indikátory jsou podle [38], str. 136, matematicky odvozené reprezentace ceny vytvořené k poskytnutí dodatečných informací, než které nám cena samotná nabízí. To, co se analytikovi nedaří vypořádat pouze z kurzů akcií, mu může připadat naprosto jasné z indikátoru, který byl z kurzů odvozen.

[27], str. 50, dále uvádí rozdělení na indikátory cenové, cenově objemové a objemové podle toho s jakou časovou řadou pracují – ceny akcie nebo objem obchodů. Navíc existují indikátory celkového trhu, které tvoří zvláštní skupinu a ukazují momentální stav trhu.

⁶⁸ „Den, kdy je možné uzavírat obchody na trhu.“ ([60])

4.3.1 Klouzavé průměry

Tato část vychází z [27], [35], [39].

Jedním z nejvíce používaných nástrojů technické analýzy, podle [27], str. 51, je indikátor založený na klouzavých průměrech. Zároveň se podle [39], str. 275, jedná o jeden z nejstarších a nejuspěšnějších nástrojů k identifikaci trendů. Patří do skupiny cenových indikátorů, tj. počítá s časovou řadou kurzů akcií – lze využít nejvyšší, nejnižší, uzavírací kurzy, denní průměry nebo jakékoliv jiné hodnoty pod podmínkou jejich konzistence po celou analyzovanou dobu. Nejčastěji jsou podle [39], str. 276, používány uzavírací kurzy akcií, které byly zvoleny i v této práci.

Jejich využití, jak [27] dále uvádí, je pro vyhlazení prudkých výkyvů dat a identifikaci rostoucího nebo klesajícího trendu. Přestože jsou klouzavé průměry oblíbenou technikou, jejich nevýhodou je zpoždění za aktuálním vývojem kurzu závislé na délce okna. Způsob použití všech indikátorů zařazených mezi klouzavé průměry je shodné.

Parametrem při použití klouzavých průměrů je délka klouzavého okna. Hodnota parametru by měla odpovídat charakteru investice a trhu. Pokud má obchodník v plánu krátkodobé investice v případě postranního trhu⁶⁹, doporučuje [27], str. 51, volit kratší klouzavé okno. V případě dlouhodobých investic na trhu s trendem (rostoucí, klesající) doporučuje volit delší klouzavé okno, aby nedocházelo k častému generování falešných signálů. Platí tedy pravidlo, „že čím je perioda kratší, tím více získáme signálů včetně signálů falešných.“ (viz [35], str. 52)

4.3.1.1 Jednoduchý klouzavý průměr

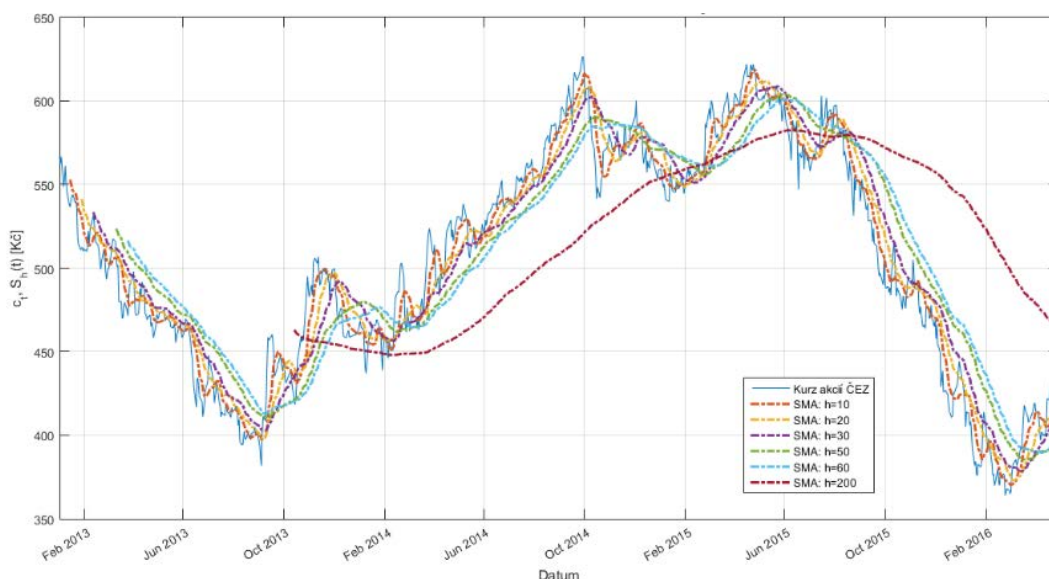
Neznámější a díky svému výpočtu nejoblíbenější (viz [35], str. 51) z této skupiny je metoda založená na jednoduchém klouzavém průměru. Myšlenka je v tomto případě podobná té, která byla uvedena v části 2.3.3 s tím rozdílem, že průměrná hodnota za období h je přiřazena na konec období namísto doprostřed, tj. vyrovnává poslední hodnotu namísto prostřední a rozsah proto nemusí být lichý. Vzorec pro jeho výpočet je (viz [27], str. 52)

$$S_h(t) = \frac{1}{h} \sum_{i=0}^{h-1} c_{t-i}, \quad \forall t \geq h, \quad (32)$$

kde $S_h(t) \in \mathbb{R}^+$ je jednoduchý klouzavý průměr délky h v čase t a
 $c_t \in \mathbb{R}^+$ je kurz akcií v čase t .

Příklad použití jednoduchého klouzavého průměru (32) na reálné kurzy akcií ČEZ očištěné o dividendy z období leden 2013 až duben 2016 (celkem 833 hodnot) ukazuje Obrázek 4.1. Délky klouzavých oken byly voleny podle nejčastěji (viz dále) volených hodnot. Jak naznačuje Obrázek 4.1, graf průměru začíná až od hodnoty rovné h , a to z důvodu jeho konstrukce – vyrovnávána je poslední hodnota v rámci daných h hodnot a k jeho celkovému výpočtu k předchozím dnům chybí data. Prvních $(h - 1)$ hodnot při výpočtu klouzavého průměru tedy chybí.

⁶⁹ Postranní trh je charakteristický postranním (horizontálním) trendem (tj. trend, který není ani rostoucí ani klesající) s malými a častými výkyvy cen, kdy je nabídka a poptávka po daném cenném papíru stále přibližně v rovnováze. (definice v [27], str. 9).



Obrázek 4.1: Jednoduché klouzavé průměry různých délek klouzavých oken spočítané z kurzu akcií ČEZ očištěného o dividendy, $n = 833$ (zdroj: vlastní simulace)

Vyhlazován je tedy původní cenový pohyb akcie, podobně jako v případě (10). Míra vyhlazení, jak uvádí [40], str. 36, opět závisí na volbě délky klouzavého okna. Podle [39], str. 278, se nejčastěji⁷⁰ volí h rovno období 200, 60, 50, 30, 20 nebo 10 dní. Při volbě delšího časového období se výpočet provádí z více pozorování, čímž je v něm zahrnuto více informace. Zahrnutím většího množství informace ale dochází k menšímu vlivu jednotlivých pozorování ve výpočtu, resp. jednotlivá pozorování se stávají méně významná. Větší odchylka pozorování proto nemusí mít velký vliv na delší klouzavý průměr. To však může být výhodou v případě, že se jedná o výjimečné, nepravidelné výkyvy v datech. Pokud však tato velká změna kurzu je počátkem zásadní změny trendu, trvá déle, než klouzavý průměr tuto změnu trendu odhalí. Z toho vyplývá, že delší klouzavý průměr identifikuje změnu trendu pomaleji, zároveň ale falešně nesignalizuje změnu trendu, když se jedná jen o krátkodobé odchylky. V případě kratšího klouzavého průměru je situace opačná. Čím kratší klouzavé okno je tedy zvoleno, tím citlivější je výsledný indikátor.

Mezi nevýhody (32), podle [35], str. 52, patří výpočet průměru v rámci omezeného časového úseku, kdy předchozí ani následující hodnoty nejsou zohledňovány. Tuto nevýhodu odstraňují modifikace jednoduchého klouzavého průměru uvedené dále.

4.3.1.2 Vážený klouzavý průměr

Vážený klouzavý průměr přihlíží ke „stáří“ každé hodnoty kurzu patřící do daného klouzavého okna. V tomto případě jsou používány lineárně rozložené váhy, které směrem do minulosti klesají, tj. největší váhu mají aktuální data a nejnižší váhu mají nejstarší data. Váhy u klouzavých průměrů se v součtu vždy rovnají jedné⁷¹. Výpočet je proveden pomocí vzorce (viz [27], str. 54)

$$W_h(t) = \frac{2}{h(h+1)} \sum_{i=0}^{h-1} (h-i)c_{t-i}, \quad \forall t \geq h, \quad (33)$$

kde $W_h(t) \in \mathbb{R}^+$ je vážený klouzavý průměr délky h v čase t ,
 $h \in \mathbb{N}$ je délka klouzavého okna a

⁷⁰ Jde o hodnoty vybírané v dobách před existencí počítačů, kdy se musely veškeré výpočty provádět ručně.

⁷¹ Platí i pro ostatní klouzavé průměry.

$c_t \in \mathbb{R}^+$ je kurz akcií v čase t .

Jednoduché i vážené klouzavé průměry jsou kritizovány za to, že jsou počítány na základě omezeného časového období. Kurzy předcházející tomuto období nejsou vůbec zohledňovány. Přesto ale podle [35], str. 53, mohou mít předchozí data určitý význam. Tuto nevýhodu odstraňuje následující typ klouzavého průměru.

4.3.1.3 Exponenciální klouzavý průměr

Exponenciální klouzavý průměr přihlíží ke všem dostupným historickým kurzům. Od váženého se liší tím, že neuvažuje váhy rozložené lineárně, jejich rozložení je v tomto případě exponenciální. Současná data mají největší váhu a směrem do minulosti váhy klesají. Vyrovnávání je v tomto případě založeno na veškerých historických datech. V tomto případě je h parametrem klouzavého průměru určující výši vyrovnávací konstanty α . Pro výpočet platí ([27], str. 55)

$$E_h(t) = \sum_{i=0}^{t-2} \alpha(1-\alpha)^i c_{t-i} + (1-\alpha)^{t-1} c_1, \quad (34)$$

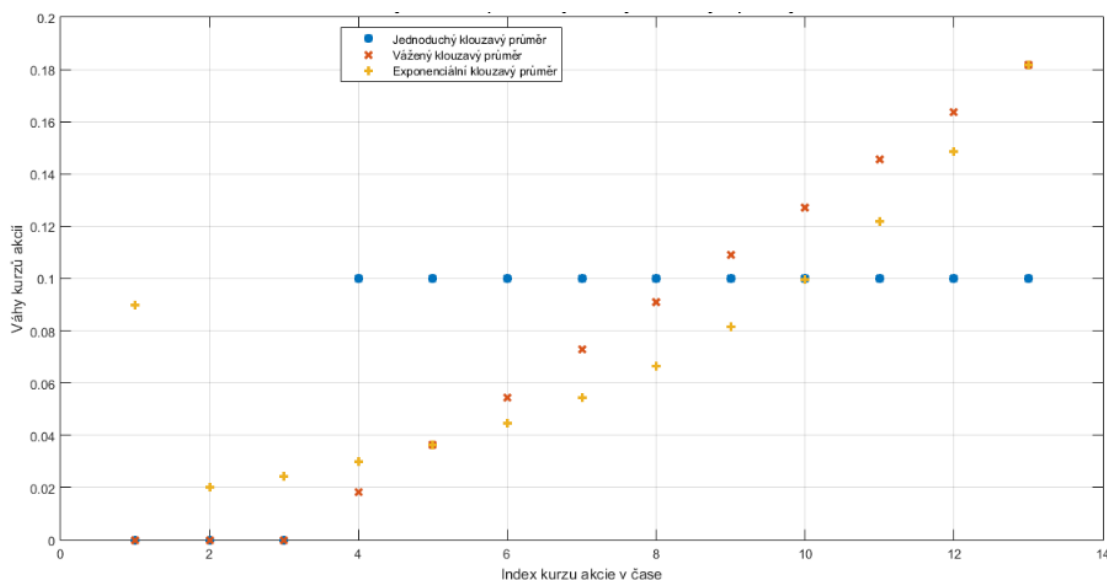
kde $E_h(t) \in \mathbb{R}^+$ je exponenciální klouzavý průměr délky h v čase t ,

$h \in \mathbb{N}$ je délka klouzavého okna,

$\alpha = \frac{2}{h+1}$ je vyrovnávací konstanta a

$c_t \in \mathbb{R}^+$ je kurz akcií v čase t .

(34) je citlivější na změny kurzu než jednoduchý klouzavý průměr ale méně citlivý než vážený (viz [27], str. 61). Obrázek 4.2 srovnává váhy používané zatím uvedenými klouzavými průměry. Délka klouzavého okna byla u všech tří průměrů zvolena deset (jedna z nejčastěji volených, viz výše). Pro demonstraci vah pozorování pro všechna dostupná pozorování (také mimo okno) bylo v tomto případě zvoleno rozpětí 13 dostupných minulých kurzů.



Obrázek 4.2: Váhy přiřazované jednotlivým kurzům akcií v čase při použití SMA, WMA a EMA, $h = 10$, $n = 13$ (zdroj: vlastní simulace)

4.3.1.4 Klouzavá regrese

Dalším možným indikátorem je klouzavá regrese. Nejedná se o další variaci klouzavého průměru. Společný je pouze „klouzavý“ způsob výpočtu. Přesto ale bývá (viz např. [27], str. 51) zařazována mezi klouzavé průměry.

V tomto případě není uvažována žádná vážená lineární kombinace několika minulých pozorování, kdy součet vah je roven jedné. Jde o prokládání přímky posledními h pozorováními metodou nejmenších čtverců (blíže viz např. [15], str. 364) stejným způsobem, jaký byl uveden v části 2.3.1.

„Nejedná se o protažení odhadované přímky o jedno pozorování dopředu (do času $t + 1$) a tedy o „předpověď“ ceny v čase $t + 1$, nýbrž se jedná o nahrazení bodu (ceny) v čase t bodem prokládané přímky v čase $t + 1$. Prokládaná přímka se přitom spočítá z cen c_t, \dots, c_{t-h+1} .“ [27], str. 60.

[27], str. 59, uvádí dva způsoby vyrovnávání. Pro tuto práci byl zvolen postup „vhodnější k výpočtům v programech, neboť podstatně zrychluje výpočet [oproti druhému postupu]“ díky konstantní regresní matici. Nechtě je tedy uvažován model

$$Y = X\beta + \varepsilon, \quad (35)$$

kde $Y = (c_{t-h+1}, c_{t-h+2}, \dots, c_{t-1}, c_t)^T$ je vektor h kurzů akcií,
 $X = \begin{pmatrix} 1, 1, \dots, 1, 1 \\ 1, 2, \dots, h-1, h \end{pmatrix}^T$ je regresní matice,
 $\beta = (\beta_t^{(0)}, \beta_t^{(1)})^T$ je vektor parametrů modelu,
 $\varepsilon = (\varepsilon_{t-h+1}, \varepsilon_{t-h+2}, \dots, \varepsilon_{t-1}, \varepsilon_t)^T$ je vektor náhodné složky modelu,
 $h \in \mathbb{N}$ je délka klouzavé regrese a
 $c_t \in \mathbb{R}^+$ je kurz akcií v čase t .

Odhad parametrů metodou nejmenších čtverců má tvar

$$b = (X^T X)^{-1} X^T Y = \begin{pmatrix} b_t^{(0)} \\ b_t^{(1)} \end{pmatrix}, \quad (36)$$

kde Y je vektor kurzů akcií z (35),
 X je regresní matice z (35),
 $b_t^{(0)} \in \mathbb{R}$ je odhad parametru $\beta_t^{(0)}$ z (35) a
 $b_t^{(1)} \in \mathbb{R}$ je odhad parametru $\beta_t^{(1)}$ z (35).

Vyrovnané kurzy akcií jsou definované jako

$$\hat{c}_t = b_t^{(0)} + b_t^{(1)}(h + 1), \quad (37)$$

kde $b_t^{(0)}, b_t^{(1)} \in \mathbb{R}$ jsou odhady definované v (36) a
 $h \in \mathbb{N}$ je délka klouzavé regrese.

Jde o indikátor odlišný od klouzavých průměrů. Extrémních hodnot (tj. lokálních maxim nebo minim) dosahuje přibližně ve stejnou dobu jako kurz akcie, případně velmi brzy poté. Extrémy odhadu (37) však jsou mnohem výraznější než u původních hodnot. „Lze říci, že klouzavá regrese „prodlužuje“ chování ceny. Naproti tomu v obdobích mírného nebo postranního trendu velmi přesně kopíruje skutečný průběh ceny a dává tedy i (příliš) mnoho signálů. Ostatní klouzavé průměry naopak výkyvy cen zmírňují, svých extrémních hodnot však dosahují s mnohem větším zpožděním, než jich dosahuje cena akcie.“ [27], str. 61.

4.3.1.5 Použití více klouzavých průměrů zároveň

Tato část vychází z [27].

V technické analýze zastávají, podle [40], str. 36, klouzavé průměry dvě funkce. V závislosti na vývoji ceny akcie indikují, jestli je býčího⁷² nebo medvědího⁷³ charakteru, a pomáhají odhadnout i její dlouhodobý charakter. Druhá funkce je pro tuto práci zajímavější. Metoda založená na klouzavém průměru generuje signály k nákupu (resp. k prodeji), které dává až poté, co cena dosáhla lokálního vrcholu (resp. dna) a již začala klesat (resp. růst) (tj. zpoždění).

Analýzu není nutné omezovat informacemi získanými z použití pouze jednoho klouzavého průměru. Pokud by bylo uvažováno současné použití více klouzavých průměrů stejného typu s různými délkami klouzavého okna, zvětšila by se množina získaných informací z vývoje kurzu akcie. Podle [27], str. 52, je tedy možné využít klouzavé průměry třemi způsoby, kdy je sledováno překřížení různých křivek. Současně lze použít klouzavý průměr a kurzy akcie, krátkodobý a dlouhodobý klouzavý průměr anebo tři klouzavé průměry s různými délkami klouzavého okna. Všechny tři způsoby jsou demonstrovány pomocí jednoduchého klouzavého průměru. U ostatních typů je postup analogický.

V případě obchodování v krátkodobém horizontu je sledováno překřížení kurzu a klouzavého průměru kurzu jako na Obrázek 4.1 nebo na Obrázek 4.3. Za použití takového indikátoru jsou signály generovány následujícím způsobem ([27], str. 52)

1. Nákup $\Leftrightarrow c_{t-1} \leq S_h(t-1) \wedge c_t > S_h(t)$ (kurz protne klouzavý průměr zdola)
 2. Prodej $\Leftrightarrow c_{t-1} \geq S_h(t-1) \wedge c_t < S_h(t)$ (kurz protne klouzavý průměr shora)
- (38)

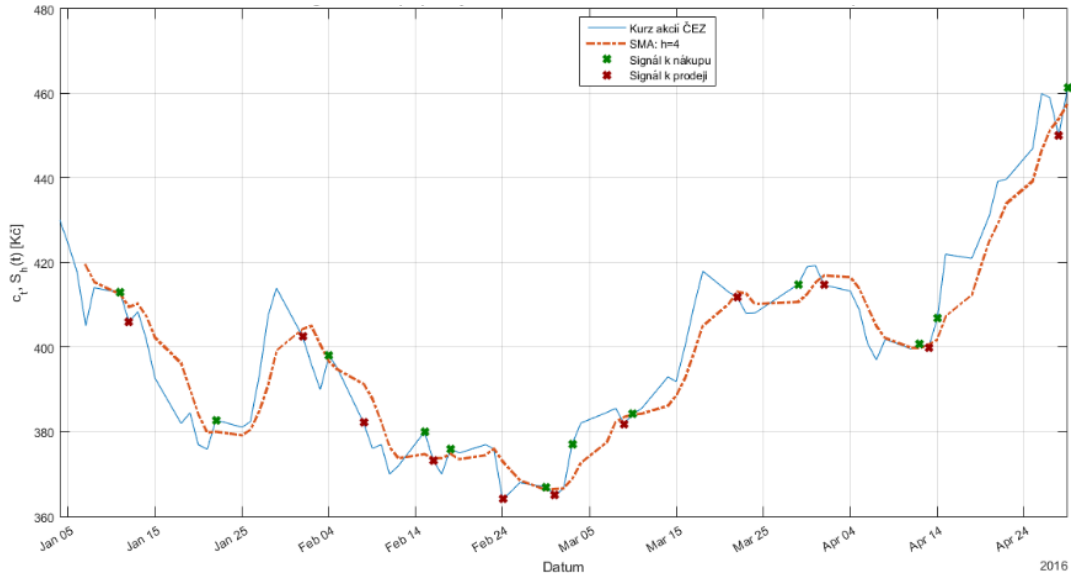
kde $c_t \in \mathbb{R}^+$ je kurz akcií v čase t a
 $S_h(t) \in \mathbb{R}^+$ je klouzavý průměr délky h v čase t .

Pravidla (38) pro generování signálů byla použita na kurzy akcií ČEZ od začátku roku 2016 do dubna (celkem 83 obchodních dní) a tyto signály byly také zaznamenány do grafu na Obrázek 4.3. Délka klouzavého okna byla zvolena na základě doporučené kombinace uvedené níže. Graf například naznačuje, že kurz měl přibližně v druhé polovině února 2016 téměř postranní trend. Na tomto

⁷² „Označení pro optimistu na akciových trzích nebo pro rostoucí trend.“ ([60]).

⁷³ „Označení pro pesimistu na akciových trzích nebo pro klesající trend.“ ([60]).

období je také vidět, že použití klouzavého průměru v případě postranního trendu může být pro investora ztrátové.



Obrázek 4.3: Generování signálů k nákupu/prodeji akcií ČEZ pomocí jednoduchého klouzavého průměru, $h = 4$, $n = 83$ (zdroj: vlastní simulace)

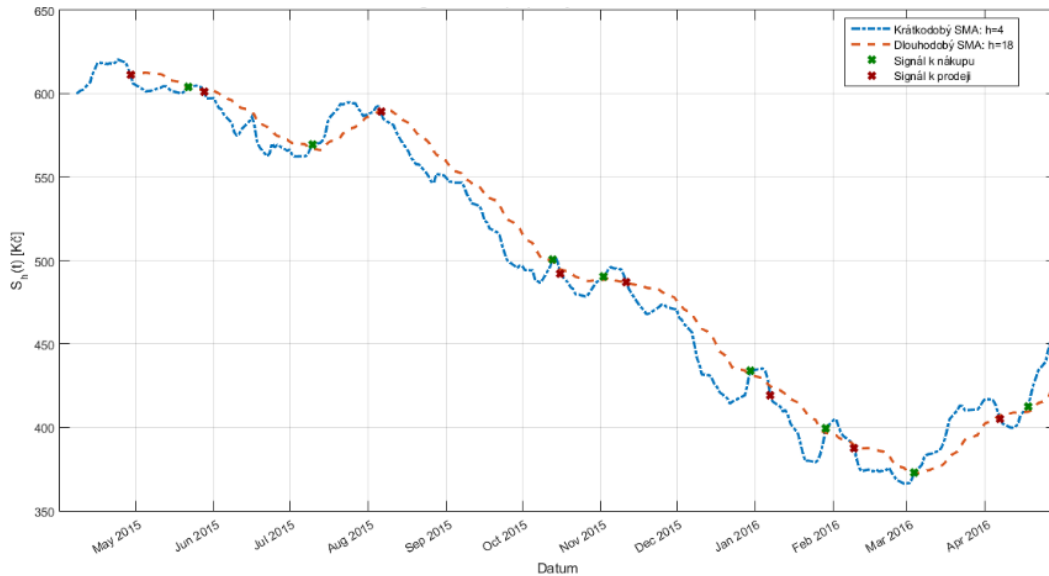
Při obchodování v dlouhodobém horizontu je sledováno překřížení krátkodobého a dlouhodobého klouzavého průměru kurzu, viz Obrázek 4.4. Za použití takového indikátoru jsou signály generovány následujícím způsobem ([27], str. 52)

1. Nákup $\Leftrightarrow S_{h_1}(t-1) \leq S_{h_2}(t-1) \wedge S_{h_1}(t) > S_{h_2}(t)$, $h_1 < h_2$,
(kurz protne klouzavý průměr zdola)
 2. Prodej $\Leftrightarrow S_{h_1}(t-1) \geq S_{h_2}(t-1) \wedge S_{h_1}(t) < S_{h_2}(t)$, $h_1 < h_2$,
(kurz protne klouzavý průměr shora)
- (39)

kde $S_{h_1}(t) \in \mathbb{R}^+$ je krátkodobý klouzavý průměr délky h_1 v čase t a
 $S_{h_2}(t) \in \mathbb{R}^+$ je dlouhodobý klouzavý průměr délky h_2 v čase t .

Pravidla (39) jsou důsledkem větší citlivosti krátkodobého průměru v porovnání s průměrem dlouhodobým. Tato pravidla pro generování signálů byla použita na přibližně jeden rok staré kurzy akcií ČEZ, tj. duben 2015 až duben 2016 (celkem 269 obchodních dní), a tyto signály byly také zaznamenány do grafu na Obrázek 4.4. Délka klouzavých oken byla pro možnost srovnání zvolena podle populární kombinace parametrů v případě použití tří klouzavých průměrů, která je uvedena níže. V grafu je vidět, že od okamžiku vygenerování signálu k prodeji 6. srpna 2015 až do okamžiku vygenerování signálu k nákupu 4. března 2016 dojde k vygenerování několika dalších „zbytečných“ signálů. V případě ignorování signálů k nákupu v tomto období, kdy kurz dlouhodobě klesal,

by došlo k celkovému snížení ztrát. Toho by mohlo být možné dosáhnout např. přidáním ještě jednoho klouzavého průměru.



Obrázek 4.4: Generování signálů k nákupu/prodeji akcií ČEZ pomocí krátkodobého ($h = 4$) a dlouhodobého ($h = 18$) jednoduchého klouzavého průměru, $n = 269$ (zdroj: vlastní simulace)

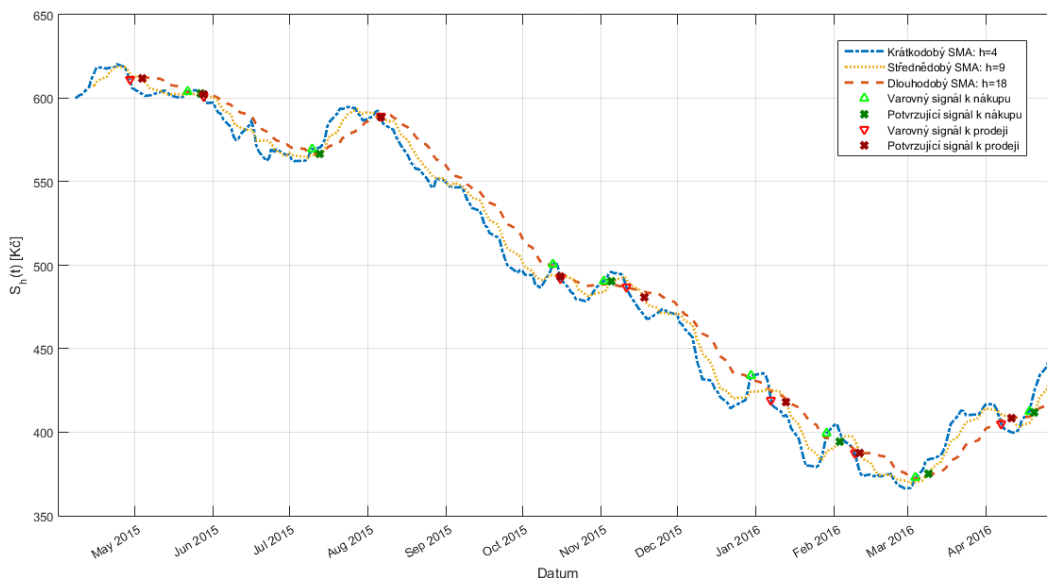
Pro větší „bezpečí“ je možné sledovat tři klouzavé průměry s různými délkami klouzavých oken. Nejčastěji používaná kombinace klouzavých průměrů podle [27], str. 52, je s parametry 4/9/18 a využívá se zejména na komoditních trzích. V takovém případě jsou generovány varující a potvrzující signály podle následujících pravidel ([27], str. 52)

1. varování $\Leftrightarrow S_{h_1}(t-1) \leq S_{h_3}(t-1) \wedge S_{h_1}(t) > S_{h_3}(t)$, $h_1 < h_2 < h_3$,
 2. nákup \Leftrightarrow potvrzení $\Leftrightarrow S_{h_2}(t-1) \leq S_{h_3}(t-1) \wedge S_{h_2}(t) > S_{h_3}(t)$,
 1. varování $\Leftrightarrow S_{h_1}(t-1) \geq S_{h_3}(t-1) \wedge S_{h_1}(t) < S_{h_3}(t)$, $h_1 < h_2 < h_3$,
 2. prodej \Leftrightarrow potvrzení $\Leftrightarrow S_{h_2}(t-1) \geq S_{h_3}(t-1) \wedge S_{h_2}(t) < S_{h_3}(t)$,
- (40)

kde $S_{h_1}(t) \in \mathbb{R}^+$ je krátkodobý klouzavý průměr délky h_1 v čase t ,
 $S_{h_2}(t) \in \mathbb{R}^+$ je střednědobý klouzavý průměr délky h_2 v čase t a
 $S_{h_3}(t) \in \mathbb{R}^+$ je dlouhodobý klouzavý průměr délky h_3 v čase t .

Pravidla (40) říkají, že varování je vygenerováno při prvním protnutí dlouhodobého průměru průměrem krátkodobým a definitivní signál k nákupu nebo prodeji při podobném protnutí průměrem střednědobým. Tento způsob využití tří jednoduchých klouzavých průměrů byl aplikován na stejná data jako v předchozím případě (tj. Obrázek 4.4) – viz Obrázek 4.5. Rozdíl mezi použitím dvou a tří klouzavých průměrů je patrný z grafu především v rámci výše diskutovaného období (tj. 6. srpna 2015 až 4. března 2016). 13. října 2015 a 30. prosince 2015 byly v tomto případě

vygenerovány pouze varovné signály k nákupu, které ale nebyly následně střednědobým průměrem potvrzeny a žádný obchod by proto nebyl uzavřen. Tím by nedošlo ke ztrátě z těchto obchodů.



Obrázek 4.5: Generování signálů k nákupu/prodeji akcií ČEZ pomocí krátkodobého ($h = 4$), střednědobého ($h = 9$) a dlouhodobého ($h = 18$) jednoduchého klouzavého průměru, $n = 269$ (zdroj: vlastní simulace)

Použitím tří různých klouzavých průměrů způsobem jako na Obrázek 4.5 by bylo dosaženo většího zisku než v případě dvou klouzavých průměrů na Obrázek 4.4 díky eliminaci alespoň některých „falešných“ signálů. Pro eliminaci co největšího počtu „falešných“ obchodů je třeba provést detailnější analýzu volby délek klouzavých oken pro jednotlivé akciové tituly.

4.3.1.6 Možná strategie obchodování

Tato část vychází z [39].

Klouzavý průměr neposkytuje žádnou konkrétní informaci o tom, jestli bude trend rostoucí nebo klesající. K profitu z použití klouzavých průměrů, je nutné, aby vývoj kurzu akcie ve sledovaném období byl „trendující“ (tj. rostoucí nebo klesající trend) a mohlo tudíž docházet k překřižování. Analytik si proto musí být jistý, že trend ve sledovaném období existuje, než začne využívat překřižování klouzavých průměrů. Přesto ale jsou někteří investoři ochotni obchodovat podle často generovaných „falešných“ signálů v případě postranního trendu, aby nepropásli okamžik změny v rostoucí (resp. klesající) trend. Velké množství chybných signálů by mohlo vést k malým ztrátám, zatímco investor čeká na jeden signál, který by vedl k velkému zisku. Tato strategie může být velice zisková v případě, že je analytik „dostatečně“ disciplinovaný a má odvahu podstupovat časté, malé ztráty. Podle [39], 281, je tato strategie často základem mnoha dlouhodobě orientovaných obchodních systémů.

Mnoho úspěšných strategií založených na klouzavých průměrech používá tyto průměry pouze pro určení trendu. Následně hledají grafické formace (viz část 4.2), podle kterých poskytují signály k nákupu nebo prodeji.

4.3.2 Další indikátory

Vzhledem k cílům této práce, jsou uvedeny pouze výše vybrané a v praktické části použité indikátory na základě jejich použitelnosti ve vztahu k jádrovým odhadům. Pro detailní popis dalších používaných indikátorů viz např. [27], str. 50-92 nebo [35], str. 51-83.

5 Data a jejich úprava

Burza poskytuje mnoho informací o vývoji kurzů akcií (otevírání, závěrečný kurz, denní maximum nebo minimum, apod.). Výpočty v této práci jsou ale založeny na historických denních závěrečných kurzech. Na vývoj kurzu má vliv vyplácení dividendy ale i štěpení akcií. Ne všechny akciové společnosti však tyto akce provádějí. Aby byla jakákoliv analýza vývoje kurzu „na místě“, je třeba pročíst zápisy zasedání valné hromady dané akciové společnosti a zjistit datum a výši vyplácených dividend a jestli a kdy společnost „štěpila“ akcie. V této kapitole jsou představeny vybrané akciové tituly, které jsou v případě potřeby očištěny od výše uvedených vlivů pomocí níže vysvětlených postupů.

5.1 Vliv dividendy na kurz a jeho eliminace

V případě, že akciová společnost vyplácí akcionářům dividendu⁷⁴, je třeba odpovídajícím způsobem upravit získané kurzy, aby byl zohledněn vliv výplaty dividend na kurz popsany níže. Podle [41] je rozhodujícím faktorem rozhodný den (resp. rozhodné datum). Pouze ten akcionář, který vlastní akcie v tento den, má právo na dividendu za předchozí rok. V tento den kurz obvykle skokově klesne. Velikost skoku přibližně odráží výši dividendy. Před provedením jakékoliv analýzy je proto žádoucí historické kurzy o dividendy očistit. K tomu využijeme návod z [42]. Podle něj počítáme tzv. „dividendové koeficienty“ založené na výši dividend jako procentní části z kurzu. Dividendový koeficient spočítáme podle vzorce

$$DK_{t_i} = 1 - \frac{d_{t_i}}{c_{t_i-1}}, \quad i = 1, 2, \dots, T, \quad (41)$$

kde $DK_{t_i} \in (0, 1)$ je dividendový koeficient založený na dividendě s ex-dividendovým datem t_i ,
 t_i je i -té ex-dividendové datum⁷⁵ v rámci celého sledovaného období,
 $d_{t_i} \in \mathbb{R}^+$ je výše dividendy k datu t_i ,
 $c_{t_i} \in \mathbb{R}^+$ je kurz akcií k datu t_i a
 $T \in \mathbb{N}$ je počet ex-dividendových dat za celé sledované období.

Dividendovým koeficientem DK_{t_i} pak vynásobíme kurzy akcie předcházející datu t_i . Očištěné kurzy tedy získáme podle vzorce (vycházíme z [42])

$$\tilde{c}_t = \begin{cases} c_t, & t = t_T, t_T + 1, \dots, n \\ DK_{t_T} \cdot c_t, & t = t_{T-1}, t_{T-1} + 1, \dots, t_T - 1 \\ DK_{t_{T-1}} \cdot DK_{t_T} \cdot c_t, & t = t_{T-2}, t_{T-2} + 1, \dots, t_{T-1} - 1 \\ \vdots & \\ \prod_{i=1}^T DK_{t_i} \cdot c_t, & t = 1, 2, \dots, t_1 - 1 \end{cases} \quad (42)$$

kde $DK_{t_i} \in (0, 1)$ je dividendový koeficient (41),
 t_i je ex-dividendové datum v období i ,

⁷⁴ „Představuje výnos pro držitele akcie definovaný jako právo na podíl na zisku společnosti.“ [60].

⁷⁵ „Označuje se tak den následující po rozhodném datu, tedy první den, kdy je již akcie obchodovaná bez práva na dividendu.“ ([41])

$c_t \in \mathbb{R}^+$	je kurz akcií k datu t ,
n	je poslední datum, ke kterému máme k dispozici kurz akcií a
$T \in \mathbb{N}$	je počet ex-dividendových dat za celé sledované období.

Příklad použití tohoto postupu na kurzy akcií ČEZ očištěné o dividendy zachycuje Obrázek 5.1.

5.2 Vliv štěpení akcií na kurz a jeho eliminace

Štěpení akcií (také split) vysvětluje a definuje [43] jako „... rozdělení všech emitovaných kmenových akcií o určitý násobek. Například pokud bude split roven 5:1, znamená to, že počet všech emitovaných akcií se zvýší pětinásobně. Současně ale také pětinásobně klesne jejich tržní kurz. Jedná se tedy pouze o technickou záležitost, která se využívá v případě, kdy je tržní kurz akcií příliš vysoký, a z toho důvodu jsou pro některé drobnější obchodníky tyto akcie obtížně obchodovatelné. Záměrem je tedy obvykle podpořit likviditu na trhu.“ Díky štěpení působí akcie na trhu „levněji“ a je tak podpořen zájem investorů o jejich nákup.

Návod z [42] uvádí tzv. „štěpící koeficient“, který je určen štěpícím poměrem

$$SK_s = \frac{1}{k}, \quad (43)$$

kde $SK_s \in (0, 1)$ je štěpící koeficient založený na štěpení akcií k datu s ,
 s je datum štěpení a
 $k \in \mathbb{N}$ je násobek zvýšení počtu akcií.

Pro příklad uvedený výše by platilo $SK_s = \frac{1}{5} = 0.2$. Štěpícím koeficientem jsou následně vynásobeny kurzy akcie předcházející datu s . K očištění historických kurzů akcie je použit vzorec

$$\tilde{c}_t = \begin{cases} c_t, & t \geq s, \\ SK_s \cdot c_t, & t < s, \end{cases} \quad (44)$$

kde $c_t \in \mathbb{R}^+$ je kurz akcií k datu t ,
 s je datum štěpení a
 $SK_s \in (0, 1)$ je štěpící koeficient (43).

5.3 Data

V této části jsou představeny akciové tituly, proces jejich výběru a také získaná data.

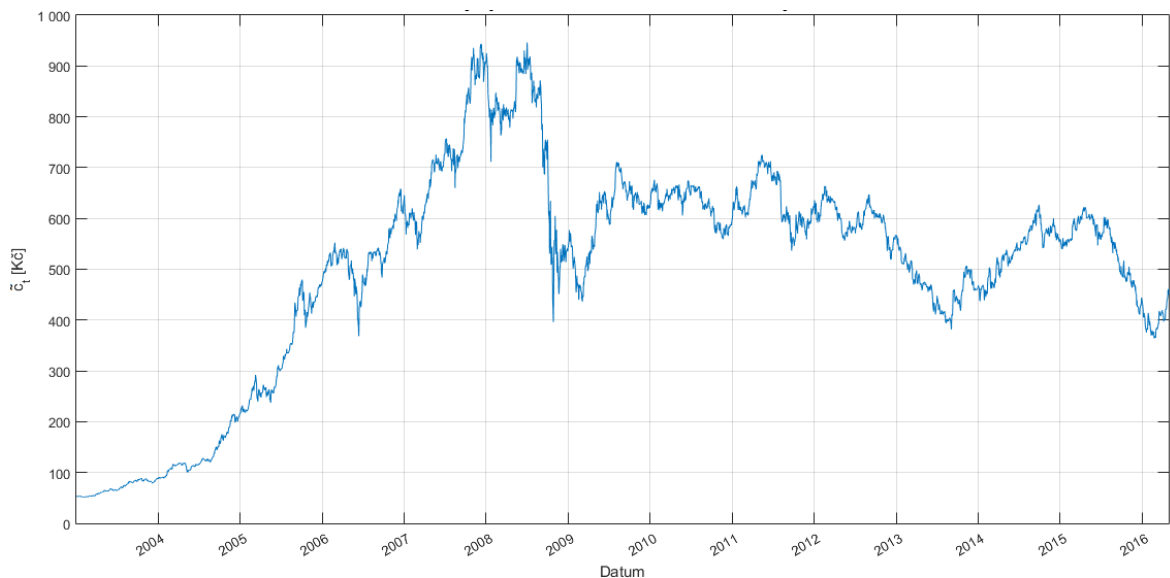
Výběrový proces akciových titulů se skládal ze tří částí. Bylo zjišťováno, které z indexu PX jsou na BCPP nejdéle a nejvíce obchodované. Následně byl výběr zúžen na tři různá hospodářská odvětví, tj. energetika, bankovní sektor a tabákové výrobky. Nakonec byl vybrán jediný zástupce každého z odvětví. Na základě tohoto výběrového procesu byly vybrány akcie společností ČEZ, a.s., Komerční banka, a.s. a Philip Morris ČR, a.s.⁷⁶ Akcie všech vybraných společností jsou obchodovány na Prime trhu BCPP. Obchodování všech těchto akciových titulů bylo zahájeno v roce 1993. Pro získaná a očištěná data viz soubor *Data.xlsx*.

⁷⁶ <http://www.cez.cz/>, <https://www.kb.cz/>, http://www.pmi.com/cs_cz/Pages/homepage.aspx

5.3.1 ČEZ, a.s.

Společnost ČEZ se zabývá výrobou, distribucí a prodejem elektrické a tepelné energie. V současné době jsou její akcie s druhou nejvyšší tržní kapitalizací⁷⁷ nejvíce⁷⁸ obchodovanou akciovou emisí na BCPP a tuto pozici v ročních souhrnech drží již od roku 2004. Zároveň tvoří z více než 23 % nejvyšší⁷⁹ podíl na tržní kapitalizaci indexu PX i PX-GLOB⁸⁰. Akcie ČEZ jsou obchodovány také v zahraničí (v Polsku a v Německu) (viz záložka „Základní informace o akcích“ v [44]). Akcionářům⁸¹ společnosti ČEZ jsou podle [44], záložka „Dividendy“, každoročně vypláceny dividendy.

Historické kurzy akcií ČEZ od ledna 2003 do dubna 2016 (tj. 3348 obchodních dní) jsme získali z [44], záložka „Vývoj cen akcií“, a z [45]. Získané kurzy byly očištěny o dividendu podle vzorce (42) a následně jejich průběh vykreslen na Obrázek 5.1.



Obrázek 5.1: Vývoj kurzu akcií ČEZ očištěného o dividendy, $n = 3348$ (zdroj dat: [44], záložka „Vývoj cen akcií“, a [45])

5.3.2 Komerční banka, a.s.

Skupina KB poskytuje komplexní služby v oblasti drobného, podnikového a investičního bankovníctví. Akcie KB jsou s třetí nejvyšší tržní kapitalizací od roku 2007 každoročně⁸² druhou nejvíce obchodovanou akciovou emisí na BCPP. S necelými 20 % tvoří třetí nejvyšší podíl na tržní kapitalizaci indexu PX i PX-GLOB. Za účelem zvýšení likvidity akcií na burzovních trzích byl na valné hromadě konané 22. 4. 2016 schválen návrh na štěpení akcií v poměru 5:1. Tento návrh, podle článku [46], nabyt platnosti dne 11. 5. 2016. Tento datum však není obsažen v žádném

⁷⁷ „Používá se jako obchodovatelná emise násobená aktuálním tržním kurzem (na burze), např. důležité akciové společnosti, popř. jako celkový objem obchodovatelných papírů na určité burze v kurzové hodnotě apod.“ [59].

⁷⁸ Viz „Tržní kapitalizace akciových emisí ke dni 23. 3. 2016“ na <http://www.pse.cz/Statistika/Trzni-Kapitalizace-Akcii-PL/> a „Nejvíce obchodované emise“ na <http://www.pse.cz/Statistika/Nejvice-Obchodovane-Emise/>.

⁷⁹ Viz „Báze indexu PX [resp. PX-GLOB] ke dni 6. 5. 2016“ na <http://www.pse.cz/Statistika/Burzovni-Indexy/default.aspx/default.aspx?bi=1>.

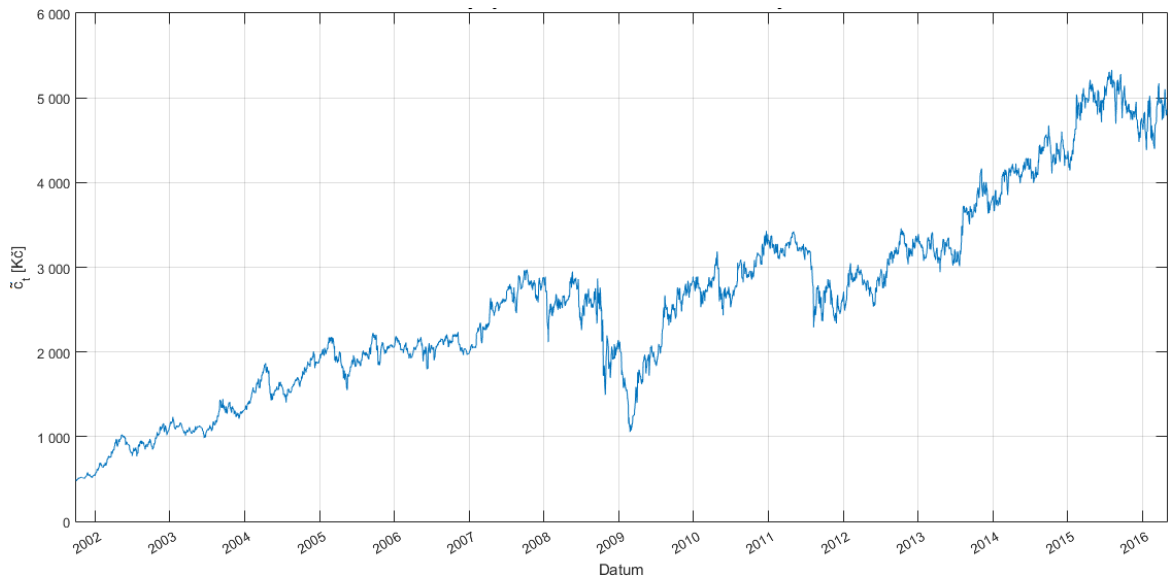
⁸⁰ Cenový index PX-GLOB je indexem BCPP se širokouází zahrnující blue chips akcie a některé další emise (blíže viz poznámky pod čarou 56 a 57).

⁸¹ „Majitel akcií, s nimiž jsou spojeny práva akcionáře [...]. Větší množství držených akcií znamená větší možnost ovlivňovat činnost akciové společnosti.“ [59].

⁸² Až na rok 2009, kdy ji vystřídaly akcie O2 Czech Republic, a.s.

z uvažovaných období, ať už pro optimalizaci parametrů modelu či pro simulaci obchodování, a proto nebyl vzorec (44) pro úpravu dat použit.

Historické kurzy akcií do dubna 2016 KB byly staženy z [47], záložka „Akcie KB“. K dispozici jsou kurzy akcií od října 2001 (tj. 3663 obchodních dní). Kurzy očištěné o dividendu podle vzorce (42) byly zaneseny na Obrázek 5.2.

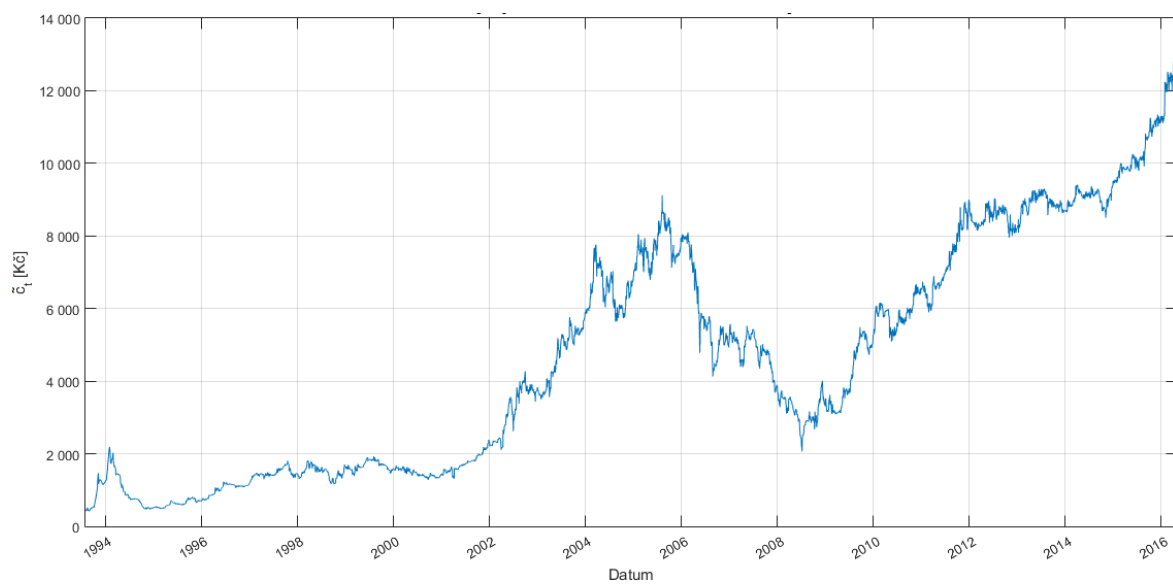


Obrázek 5.2: Vývoj kurzu akcií KB očištěného o dividendy, $n = 3663$ (zdroj dat: [47], záložka „Akcie KB“)

5.3.3 Philip Morris ČR, a.s.

Společnost Philip Morris ČR, a.s. spadá pod mezinárodní skupinu Philip Morris International, Inc. a je největším výrobcem a prodejcem tabákových výrobků v ČR. Na českém trhu působí již od roku 1992. Obchodování s akciemi PM bylo na BCPP zahájeno v červenci roku 1993. Na tržní kapitalizaci indexu PX se podílí téměř 4 %. Z hlediska tržní kapitalizaci jí patří osmé místo v rámci všech akciových titulů obchodovaných na BCPP. V posledních osmi letech jsou její akcie mezi devíti nejvíce obchodovanými emisemi. Akcionářům jsou podle [48], záložka „Dividendy“, každoročně vypláceny dividendy. Roku 2014 bylo na valné hromadě rozhodnuto o zrušení rezervního fondu společnosti a její představenstvo rozhodlo o rozdělení zůstatku ve formě vyplacení dividendy. Vypořádání se s touto mimořádnou událostí nemá na úpravu dat žádný další vliv. Byla tedy pouze zvýšena dividendy o odpovídající částku.

Historické kurzy akcií PM do dubna 2016 byly staženy z [49]. K dispozici jsou kurzy od srpna 1993 (tj. 5521 obchodních dní). Kurzy očištěné o dividendu a mimořádnou výplatu akcionářům podle vzorce (42) byly zaneseny na Obrázek 5.3.



Obrázek 5.3: Vývoj kurzu akcií PM očištěného o dividendy, $n = 5521$ (zdroj dat: [49])

6 Obchodní systém

Pro obchodování akciových titulů na BCPP byl vytvořen obchodní systém složený ze dvou částí. V první části je provedena optimalizace parametrů na „trénovacích“ datech a v druhé části je simulováno obchodování s optimalizovanými parametry. Pro rozsah trénovacích dat byla zvolena podmínka, aby uzavřených obchodů bylo alespoň deset (u každého modelu při doporučené kombinaci parametrů) a zároveň byla použita data za celý kalendářní rok. Obchodování je simulováno na datech vždy od ledna 2014 do dubna 2016. Simulační studie je provedena vždy pro jeden akciový titul. Parametry jsou tedy vybírány „na míru“. Výsledky získané pomocí vytvořeného obchodního systému jsou srovnávány s jednou z naivních strategií obchodování (viz část 6.3.1).

6.1 Popis navržených modelů

Ve vytvořeném obchodním systému je možné volit mezi osmi modely, kterými jsou čtyři „klouzavé“ modely (SMA, WMA, EMA a klouzavá regrese) podle technické analýzy a čtyři modely založené na jádrových odhadech (možnost volby mezi čtyřmi různými jádrovými funkcemi). Ve všech případech je pro obchodování použit stejný systém dvou typů signálů, varovný a potvrzující, které jsou generovány podle pravidel (40). Tento systém byl zvolen na základě analýzy použití klouzavých průměrů v části 4.3.1.5. Ke správné funkci takového obchodního systému je proto potřeba mít k dispozici tři křivky s různými parametry.

Každý model zohledňuje také poplatky za uzavírání burzovních obchodů, uvedené v Tabulka 2. Daň z příjmu v případě splnění podmínek uvedených v podkapitole 3.2 není v modelech uvažována.

6.1.1 Modely založené na klouzavých průměrech

Modely založené na klouzavých průměrech využívají vždy tři klouzavých průměrů stejného typu s různými šířkami vyhlazovacích oken. Tyto modely jsou popsány v částech 4.3.1.1 až 4.3.1.3. Stejným způsobem je použit i model založený na klouzavé regresi popsáný v části 4.3.1.4. Všechny jsou dále používány způsobem popsáným v části 4.3.1.5, viz Obrázek 4.5. Modely založené na klouzavých průměrech jsou dále označovány podle názvů z angličtiny, tj. „SMA“, „WMA“ a „EMA“.

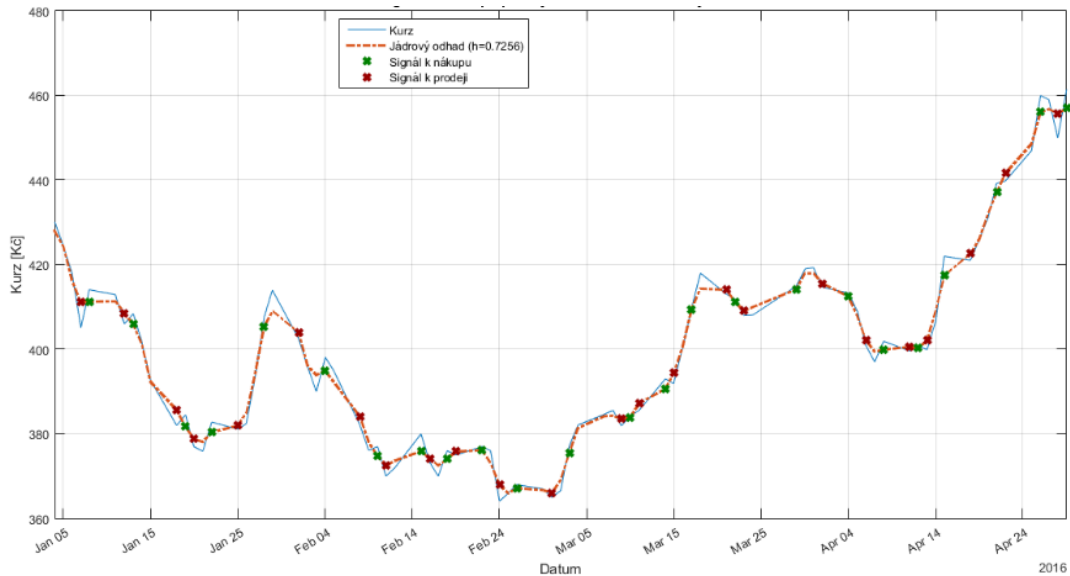
6.1.2 Modely založené na jádrových odhadech

Sloučením teorie jádrových odhadů a technické analýzy byl vytvořen vlastní „indikátor“. Ten využívá Nadaraya-Watsonův jádrový odhad (22), popsáný v části 2.3.4.1, a pravidla (40) pro použití tří klouzavých průměrů zároveň. Je vycházeno z myšlenky, že jádrové odhady regresní křivky jsou ve své podstatě obdobou klouzavých průměrů. Hlavním rozdílem se zdá být volba vah pro jednotlivá pozorování, tj. kurzy v jednotlivých dnech.

6.1.2.1 Použití jádrového odhadu

K tomu, aby bylo možné použít jádrové odhady i pro obchodování s akciemi, je třeba upravit použití vzorce (22) podobným způsobem jako technická analýza upravuje definici klouzavých průměrů, tj. přiřazení odhadované hodnoty na konec okna místo doprostřed. Pokud by nebyl tento posun

zohledněn, bylo by systémem vygenerováno příliš velké množství signálů a stal by se v praxi nepoužitelným. Příklad takového systému je uveden na Obrázek 6.1. Pro tento obrázek byla zvolena stejná data jako pro Obrázek 4.3, tj. kurzy akcií ČEZ očištěné o dividendy od začátku roku 2016 do konce dubna 2016 (celkem 83 obchodních dní). Epanechnikova jádrová funkce byla zvolena na základě výsledků procesu optimalizace pro akcie ČEZ, kterých bylo dosaženo v části 6.4.1.



Obrázek 6.1: Obchodní systém pro akcie ČEZ používající jádrový odhad s Epanechnikovou jádrovou funkcí, odhad parametru h metodou křížového ověřování, $\hat{h} = 0.7256$, $n = 83$ (zdroj: vlastní simulace)

Pojetí jádrového odhadu se v tomto případě liší od „klasického“ odhadu regresní křivky. Rozdíl je v účelu použití tohoto odhadu. Jak lze vidět také na Obrázek 6.1, zatímco „klasický“ odhad aproximuje „co nej přesněji“ danou křivku, pro obchodování je důležité „co nej přesněji“ identifikovat okamžiky „zratu“ ve vývoji kurzu. Toho je možné dosáhnout metodou „zpoždění“ odhadu používanou v technické analýze.

6.1.2.2 „Zpožděný“ jádrový odhad

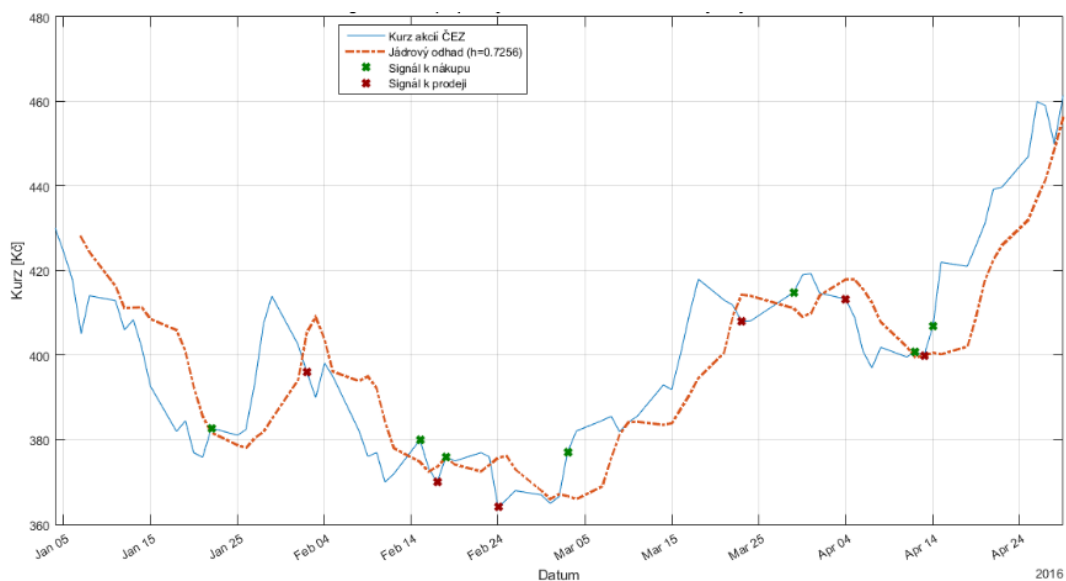
Jádrový odhad tedy bude prováděn podle vzorce (22), výsledné hodnoty ale budou „posunuté“. Posunutí nelze volit na základě „optimalizovaného“ \hat{h} , jelikož se nemusí jednat o celé číslo. Je však možné použít nový parametr „posunu“ \tilde{h} . Celá křivka je tedy posunuta o $(\tilde{h} - 1)$ dní dopředu tak, aby první odhad byl dostupný ke dni \tilde{h} . Tím vznikne zpoždění podobné tomu u modelu „Průměr“.

Posunutí je definováno následujícím způsobem. Necht' $x_i, i = 1, 2, \dots, n$, jsou burzovní dny, pro které jsou k dispozici odpovídající závěrečné kurzy akcií y_i . Společně tvoří tato data uspořádané dvojice $\{(x_i, y_i)\}_{i=1}^n$, kde n je počet dostupných dat. Dále je metodou křížového ověřování proveden odhad vyhlazovacího parametru $\hat{h} \in \mathbb{R}^+$ a existuje parametr posunu $\tilde{h} \in \mathbb{N}$, libovolný. Jádrový odhad je vypočítán pro dny $x_1, x_2, \dots, x_{n-(\tilde{h}-1)}$. Hodnoty jsou ale vlivem parametru posunutí přiřazeny dnům $x_{1+(\tilde{h}-1)} = x_{\tilde{h}}, x_{2+(\tilde{h}-1)} = x_{\tilde{h}+1}, \dots, x_{n-(\tilde{h}-1)+(\tilde{h}-1)} = x_n$, tj. vzniknou uspořádané dvojice odhadů $\left\{ \left(x_{i+(\tilde{h}-1)}, \hat{m}_{\hat{h}}(x_i) \right) \right\}_{i=1}^n$, kde $\hat{m}_{\hat{h}}(x_i)$ je hodnota Nadaraya-Watsonova jádrového odhadu pro den x_i .

Z postupu je patrné, že odhad v prvních $(\tilde{h} - 1)$ dnech, stejně jako v případě klouzavých průměrů, „chybí“. Pro posledních $n - (n - (\tilde{h} - 1)) = \tilde{h} - 1$ dní sice jsou kurzy známy, odhady v nich

ale nejsou provedeny. Z konstrukce jádrového odhadu však vyplývá, že je vždy počítáno se všemi dostupnými kurzy $y_i, i = 1, 2, \dots, n$.

Parametr posunu lze volit podle doporučení technické analýzy zmíněného v části 4.3.1.5, tj. kombinace 4/9/18. Použití jádrového odhadu v obchodním systému je analogické použití jednoho klouzavého průměru zároveň s kurzy akcií, viz Obrázek 4.3. Kvůli velké oscilaci kurzu v porovnání s odhadnutou křivkou je takový obchodní systém definovaný podmínkami generování příkazů (38) vhodný pro obchodování v krátkodobém horizontu. Pro parametr posunu se tedy doporučuje volit hodnotu 4. Příklad použití obchodního systému založeného na jádrovém odhadu vývoje kurzu akcie ČEZ s vypočteným „vhodným“ vyhlazovacím parametrem podle metody křížového ověřování je zachycen na Obrázek 6.2. Jsou použita stejná data jako pro Obrázek 6.1. Takový systém, z dlouhodobého hlediska, generuje příliš mnoho signálů a mohl by se stát velmi ztrátovým.



Obrázek 6.2: Obchodní systém pro akcie ČEZ používající jádrový odhad s Epanechnikovou jádrovou funkcí (krátkodobý horizont), odhad parametru h metodou křížového ověřování, $\tilde{h} = 4$, $n = 83$ (zdroj: vlastní simulace)

6.1.2.3 Odhad vyhlazovacího parametru a parametru posunu

Tato práce se však zaměřuje na dlouhodobé obchodování s akciemi. V takovém případě je pro „bezpečnou“ identifikaci bodů zvratu „vhodné“ použít tři různé odhady vývoje kurzu zohledňující tři různé časové horizonty, tj. krátkodobý, střednědobý a dlouhodobý (viz část 4.3.1.5). Pro použití tří zpožděných jádrových odhadů je třeba „vhodně“ zvolit tři různé vyhlazovací parametry. Metoda křížového ověřování popsaná v části 2.3.4.3 však poskytuje pouze jeden „vhodný“ vyhlazovací parametr ve smyslu minimalizace funkce křížového ověřování (30). Pro další postup proto není její použití vhodné, jelikož nedokáže zohlednit různé investiční horizonty.

Nechť proto nadále platí předpoklad, že parametr posunu jádrového odhadu kurzu je zároveň odhadem vyhlazovacího parametru, tj. $\hat{h} = \tilde{h}$ (dále označovány také jen jako parametry modelu), a platí značení \tilde{h}_K pro parametr krátkodobého horizontu, \tilde{h}_S pro parametr střednědobého horizontu a \tilde{h}_D pro parametr dlouhodobého horizontu. Hodnoty těchto parametrů mohou být uváděné v notaci $\tilde{h}_K/\tilde{h}_S/\tilde{h}_D$.

Kombinaci těchto parametrů je možné „optimalizovat“ numericky. K tomu je nutné zvolit startovací hodnoty, za které je možné volit např. kombinaci doporučenou technickou analýzou (4/9/18). „Optimalizaci“ kombinace parametrů modelu se blíže věnuje podkapitola 6.2.

6.1.2.4 Jednotlivé modely

V obchodním systému založeném na jádrovém modelu lze použít libovolnou jádrovou funkci splňující podmínky (2). V této práci jsou uvažovány funkce uvedené v Tabulka 1. Jádrové modely jsou dále rozlišované podle použité jádrové funkce a označované „Parzen“ (model používající Parzenovu jádrovou funkci), „Epanechnik“ (model používající Epanechnikovu jádrovou funkci), „Trojúhelník“ (model používající trojúhelníkovou jádrovou funkci) a „Gauss“ (model používající Gaussovu jádrovou funkci). Volbě „vhodné“ jádrové funkce se blíže věnuje podkapitola 6.2.

6.2 Optimalizace parametrů a modelu

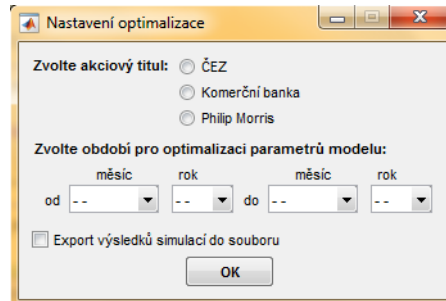
Přestože jsou vytvořeny jednotlivé modely, je třeba se nejprve zabývat tím, jaký model a s jakými parametry je pro obchodování jednotlivých akciových titulů „nejlepší“. Jak již bylo zmíněno v části 6.1.2.3, „nejlepší“ model a kombinace parametrů pro obchodování bude určována numericky.

K tomu, aby byl vybrán „nejlepší“ model numerickou cestou, musí být dopředu zvoleno, mezi jakými modely se bude vybírat. Modely podle použité váhové funkce jsou definované v části 6.1.2.4. Zbývá tedy omezit volbu parametrů jednotlivých křivek. Jak již bylo dříve diskutováno, za výchozí kombinaci lze považovat doporučení technické analýzy, tj. 4/9/18. Pro určení „vhodné“ kombinace parametrů se provede simulace obchodování na trénovacích datech. Pro omezení volby jednotlivých parametrů lze vyjít z poměru mezi doporučenými parametry, který je $\frac{\tilde{h}_S}{\tilde{h}_D} = \frac{9}{18} = 50\%$ a $\frac{\tilde{h}_K}{\tilde{h}_S} = \frac{4}{9} \doteq 44.44\%$. Necht' je zavedeno značení testovaných množin \tilde{H}_D pro rozpětí \tilde{h}_D , \tilde{H}_S pro rozpětí \tilde{h}_S a \tilde{H}_K pro rozpětí \tilde{h}_K . Testovaná rozpětí pro každý parametr má smysl volit tak, aby nedošlo k překrytí jejich hodnot, jelikož by se např. parametr dlouhodobého horizontu stal parametrem střednědobého horizontu, tj. došlo by k záměně jejich původního významu. Pro \tilde{h}_D tedy lze volit rozpětí ± 6 obchodních dní, tj. $\tilde{H}_D = \{12, 13, \dots, 24\}$. Tomu odpovídá rozpětí pro \tilde{h}_S : $\pm 50\% \cdot 6 = \pm 3$ obchodní dny, tj. $\tilde{H}_S = \{6, 7, \dots, 12\}$, a pro rozpětí \tilde{h}_K platí $\pm 44.44\% \cdot 3 \doteq \pm 1$ obchodní den, tj. $\tilde{H}_K = \{3, 4, 5\}$. Zápis v notaci uvedené výše je $4 \pm 1/9 \pm 3/18 \pm 6$. Přibližný 50% poměr zůstává zachován v obou případech. Zaokrouhlování v případě \tilde{h}_K nevnáší do procesu optimalizace chybu, jelikož se jedná o nastavení omezení startovacích hodnot a otestovány budou všechny povolené kombinace. Jiné rozsahy zřejmě nemá cenu uvažovat, jelikož by v případě jejich překrytí došlo k záměně např. parametru pro dlouhodobý odhad za parametr pro střednědobý odhad. Z výpisu jednotlivých množin je patrné, že platí $\tilde{H}_K \cap \tilde{H}_S = \emptyset$, $\tilde{H}_K \cap \tilde{H}_D = \emptyset$ a $\tilde{H}_S \cap \tilde{H}_D = \{12\}$. Tento problém lze vyřešit tak, že v případě, že jeden z parametrů nabývá hodnoty 12, nemá smysl, aby druhý parametr nabýval stejné hodnoty, jelikož obě křivky by v tom případě byly shodné, a systém tudíž nepoužitelný. Celkový počet povolených a testovaných kombinací proto je $(2 \cdot 1 + 1)(2 \cdot 3 + 1)(2 \cdot 6 + 1) - 3 = 270$.

6.2.1 Princip „optimalizace“

Proces „optimalizace“ modelu je vyřešen v příloženém skriptu *Optimalizace.m*. Po jeho spuštění je zavolán pomocný skript *nastaveni_optimalizace.m*. Ten zobrazí menu (viz Obrázek 6.3), ve kterém uživatel volí, pro který akciový titul a na základě jakého období chce „optimalizaci“ provést. V této práci je období u každého akciového titulu voleno na základě podmínek stanovených v úvodu

kapitoly 6. Vývoj kurzu akcií ČEZ splňuje podmínky v období od roku 2009 do roku 2013 (minimálně 11 obchodů), tj. 5 let (1257 obchodních dní), akcie KB je splňuje od roku 2010 do roku 2013 (minimálně 12 obchodů), tj. 4 roky (1006 obchodních dní), a PM je splňuje od roku 2009 do roku 2013 (minimálně 11 obchodů), tj. 5 let (1256 obchodních dní⁸³). Počet uzavřených obchodů podle vytvořeného obchodního systému je testován skriptem *generuj_signaly.m*, který generuje signály k nákupu a prodeji.



Obrázek 6.3: Menu pro nastavení optimalizace (zdroj: vlastní implementace)

Následně je provedena simulace obchodování pro každou z povolených kombinací. Pro potřeby této analýzy je pojem obchodování chápán jako procházení zvoleného období a testování splnění podmínek pro generování signálů k nákupu (resp. prodeji) akcií. Jednotlivé modely jsou srovnávány na základě relativního zisku (resp. ztráty) z každého uzavřeného obchodu (dlouhé pozice – viz část 6.3.1), tj.

$$z_i^{[+/-]} = \frac{\tilde{c}_{P_i}}{\tilde{c}_{N_i}} - 1, \quad i = 1, 2, \dots, I, \quad (45)$$

kde $z_i^{[+/-]} \in \mathbb{R}$ je relativní zisk (resp. ztráta) z i -tého uzavřeného obchodu,
 $\tilde{c}_{P_i} \in \mathbb{R}^+$ je kurz, za který jsou akcie v i -tém uzavřeném obchodě prodány,
 $\tilde{c}_{N_i} \in \mathbb{R}^+$ je kurz, za který jsou akcie v i -tém uzavřeném obchodě nakoupeny a
 $I \in \mathbb{N}$ je počet uzavřených obchodů v předem zvoleném období.

Skript pro výběr „vhodného“ modelu je řešen pomocí čtyř *for*-cyklů, kdy první tři cykly prochází každý jednu testovanou množinu \tilde{H}_D , \tilde{H}_S a \tilde{H}_K . Čtvrtý cyklus prochází možné modely popsané v podkapitole 6.1 s nastavenými parametry. V každé iteraci je tedy simulováno obchodování zvoleného akciového titulu a zaznamenává se zisk (resp. ztráta) v podobě vzorce (45) z každého provedeného obchodu. Pro představu o průběhu každé iterace čtvrtého cyklu a také pro srovnání jednotlivých modelů byla vytvořena Tabulka 3. Ta srovnává výsledky v rámci prvních čtyř obchodů při doporučené kombinaci parametrů 4/9/18. Z Tabulka 3 je patrné, že identifikace „vhodných“ okamžiků k nákupu (resp. prodeji) akcií závisí i na volbě modelu, tj. v případě jádrových odhadů i na volbě jádrové funkce.

⁸³ Kurz PM ze dne 29. 3. 2013 chybí.

Model	1. obchod		2. obchod		3. obchod		4. obchod	
	Datum nákup	Datum prodej	Datum nákup	Datum prodej	Datum nákup	Datum prodej	Datum nákup	Datum prodej
	Kurz [Kč]	Kurz [Kč]	Kurz [Kč]	Kurz [Kč]	Kurz [Kč]	Kurz [Kč]	Kurz [Kč]	Kurz [Kč]
	$z_1^{[+/-]}$		$z_2^{[+/-]}$		$z_3^{[+/-]}$		$z_4^{[+/-]}$	
SMA	16. 3. 2009	23. 6. 2009	15. 7. 2009	24. 8. 2009	6. 10. 2009	15. 10. 2009	2. 11. 2009	13. 11. 2009
	484.67,-	597.26,-	642.49,-	699.92,-	664.02,-	622.75,-	636.75,-	627.05,-
	+ 23.23 %		+ 8.94 %		- 6.22 %		- 1.52 %	
WMA	12. 3. 2009	19. 6. 2009	13. 7. 2009	20. 8. 2009	1. 10. 2009	14. 10. 2009	29. 10. 2009	11. 11. 2009
	470.35,-	637.46,-	620.95,-	692.02,-	673.36,-	644.64,-	649.67,-	640.33,-
	+ 35.53 %		+ 11.45 %		- 4.26 %		- 1.44 %	
EMA	13. 3. 2009	23. 6. 2009	15. 7. 2009	31. 8. 2009	28. 12. 2009	5. 2. 2010	9. 3. 2010	6. 5. 2010
	485.43,-	597.26,-	642.49,-	666.18,-	625.98,-	617.36,-	639.62,-	639.04,-
	+ 23.04 %		+ 3.69 %		- 1.38 %		- 0.09 %	
Klouzavá regrese	11. 2. 2009	3. 3. 2009	11. 3. 2009	26. 3. 2009	9. 4. 2009	22. 4. 2009	4. 5. 2009	15. 5. 2009
	474.17,-	438.91,-	475.81,-	524.92,-	536.86,-	539.25,-	582.91,-	616.72,-
	- 7.44 %		+ 10.32 %		+ 0.44 %		+ 5.80 %	
Parzen	2. 4. 2009	9. 7. 2009	3. 8. 2009	18. 9. 2009	13. 1. 2010	23. 2. 2010	31. 3. 2010	20. 5. 2010
	528.67,-	613.77,-	703.51,-	651.82,-	655.77,-	624.47,-	637.46,-	638.90,-
	+ 16.10 %		- 7.35 %		- 4.77 %		+ 0.23 %	
Epanechnik	31. 3. 2009	7. 7. 2009	31. 7. 2009	18. 9. 2009	12. 1. 2010	22. 2. 2010	30. 3. 2010	20. 5. 2010
	501.39,-	587.93,-	693.46,-	651.82,-	652.83,-	630.28,-	646.08,-	638.90,-
	+ 17.26 %		- 6.00 %		- 3.45 %		- 1.11 %	
Trojúhelník	30. 3. 2009	7. 7. 2009	29. 7. 2009	18. 9. 2009	11. 1. 2010	18. 2. 2010	29. 3. 2010	20. 5. 2010
	496.48,-	587.93,-	659.72,-	651.82,-	659.79,-	621.02,-	642.42,-	638.90,-
	+ 18.42 %		- 1.20 %		- 5.88 %		- 0.55 %	
Gauss	30. 3. 2009	3. 7. 2009	29. 7. 2009	16. 9. 2009	11. 1. 2010	18. 2. 2010	25. 3. 2010	20. 5. 2010
	496.48,-	589.15,-	659.72,-	656.13,-	659.79,-	621.02,-	641.05,-	638.90,-
	+ 18.67 %		- 0.54 %		- 5.88 %		- 0.34 %	

Tabulka 3: Data pro první čtyři doporučené obchody s akciemi ČEZ od ledna 2009 do prosince 2013, srovnání jednotlivých modelů s parametry 4/9/18 (zdroj: vlastní simulace)

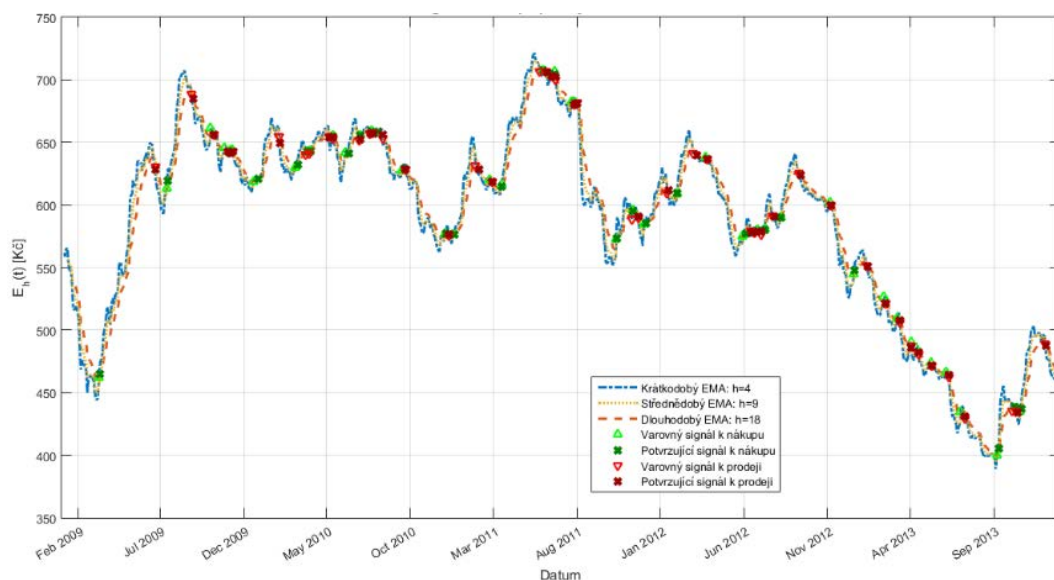
Pro představu o informacích získaných po ukončení všech iterací čtvrtého cyklu byla vytvořena Tabulka 4. Zde jsou shrnuty výsledky obchodování akcií ČEZ v rámci celého období zvoleného pro optimalizaci modelu s parametry 4/9/18. Díky charakteru vztahu (45) je u modelů porovnávána jejich schopnost identifikovat „vhodné“ okamžiky nákupu a prodeje akcií.

Model	Medián $z_i^{[+/-]}$	Průměr $z_i^{[+/-]}$	Počet vygenerovaných obchodů
SMA	- 0.60 %	+ 0.27 %	36
WMA	- 1.00 %	+ 0.64 %	40
EMA	- 1.01 %	+ 0.94 %	24
Klouzavá regrese	- 0.98 %	+ 0.12 %	67
Parzen	- 0.19 %	+ 0.78 %	11
Epanechnik	+ 0.12 %	+ 1.30 %	11
Trojúhelník	- 0.28 %	+ 1.85 %	11
Gauss	+ 1.24 %	+ 1.94 %	11

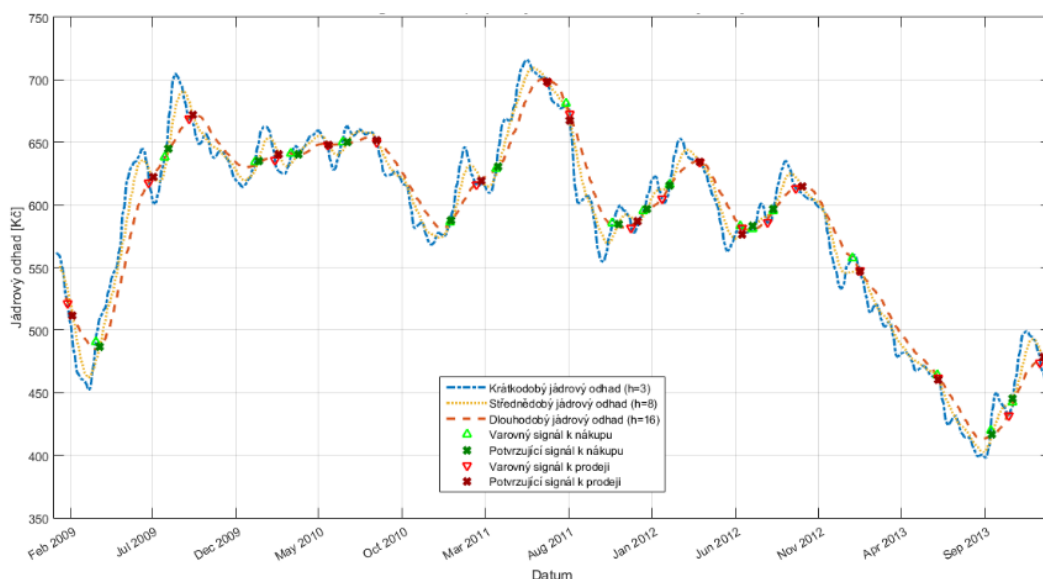
Tabulka 4: Výsledky doporučeného obchodování s akciemi ČEZ v rámci testovacího období, srovnání jednotlivých modelů s parametry 4/9/18 (zdroj: vlastní simulace)

Srovnání obchodních systémů založených na modelech EMA, používaným technickou analýzou, který se na základě definovaných podmínek ukázal být nejméně vhodným, a Gauss, který se ukázal být „optimálním“ modelem pro obchodování akcií ČEZ při doporučeném nastavení parametrů 4/9/18 zachycují Obrázek 6.4 a Obrázek 6.5. Model EMA generuje mnohem více signálů. Na každý obchod se navíc váží poplatky za každý nákup a za každý prodej. Z toho důvodu se může jevit jako velice nevýhodný v porovnání s ostatními uvažovanými modely. Tabulka 4 navíc ukazuje, že volba jádrové funkce v této oblasti aplikace jádrových odhadů je významná, jak bylo také naznačeno pomocí Tabulka 3.

Po ukončení čtvrtého cyklu je vybrán maximální medián $z_i^{[+/-]}$ ze všech modelů, tj. maximum z druhého sloupce Tabulka 4. Medián byl zvolen jako kritérium z důvodu jeho robustnosti (v porovnání například s aritmetickým průměrem). Tento maximální medián a odpovídající model jsou pro dané nastavení parametrů uloženy do tabulky v příloženém souboru *Optimalizace-výstup.xlsx*. Tabulka 5 naznačuje, jak vypadá výsledná tabulka pro výběr „optimálního“ modelu a parametrů s doplněnou hodnotou na základě Tabulka 4.



Obrázek 6.4: Obchodní systém pro akcie ČEZ aplikovaný na testovací data od ledna 2009 do prosince 2013 používající exponenciální klouzavé průměry, kombinace parametrů 4/9/18 (zdroj: vlastní simulace)



Obrázek 6.5: Obchodní systém pro akcie ČEZ aplikovaný na testovací data od ledna 2009 do prosince 2013 používající jádrový odhad s Gaussovou jádrovou funkcí, kombinace parametrů 4/9/18 (zdroj: vlastní simulace)

Kombinace parametrů			Optimální model s maximálním mediánem $z_i^{[+/-]}$	Medián $z_i^{[+/-]}$
\tilde{h}_K	\tilde{h}_S	\tilde{h}_D		
...
4	9	18	Gauss	+ 1.24 %
...

Tabulka 5: Příklad vložení záznamu o optimálním modelu pro danou kombinaci parametrů do tabulky (zdroj: vlastní soubor *Optimalizace-výstup.xlsx*)

Tento proces je tedy proveden 270-krát (viz výše). Výběr „vhodného“ nastavení pro daný akciový titul již probíhá v rámci všech těchto simulací. Opět se vybere maximální medián, tj. maximum ze všech maximálních mediánů. Tato hodnota tak určí „optimální“ kombinaci parametrů a „optimální“ model pro daný akciový titul. Kompletní tabulky výsledků všech simulací pro všechny tři akciové tituly lze nalézt v souboru *Optimalizace-výstup.xlsx* rozřazené podle jednotlivých akciových titulů.

6.2.2 Optimální nastavení pro jednotlivé akciové tituly

Z důvodu rozsáhlosti kompletních tabulek pro jednotlivé akciové tituly, vytvořených podle vzoru Tabulka 5, je zde uveden pouze postup k jejich dosažení a výsledné „optimální“ nastavení parametrů. Všechny akciové tituly byly testovány na období splňujícím podmínky z úvodu kapitoly 6. Tabulka 6 shrnuje výsledky procesu „optimalizace“. Při analýze ČEZ a KB bylo nalezeno více shodných maximálních hodnot mediánu. Pouze v případě PM vzešel z tohoto procesu jediný „optimální“ model. Tečka (•) v tabulce značí libovolnou hodnotu z odpovídající množiny, ze které je parametr volen.

Akciový titul	Kombinace parametrů			Model	Medián $z_i^{[+/-]}$
	\tilde{h}_K	\tilde{h}_S	\tilde{h}_D		
ČEZ	•	8	17	Epanechnik	+ 2.32 %
	3	8	16		
	3	9	15		
	•	9	16	Trojúhelník	
	•	10	16	Gauss	
	3	11	14		
3	12	13			
KB	•	10	12	Trojúhelník	+ 1.95 %
PM	3	10	20	Epanechnik	+ 9.72 %

Tabulka 6: „Optimální“ obchodní modely pro jednotlivé akciové tituly (zdroj: vlastní simulace)

6.2.3 Popis optimalizačního algoritmu

Proces optimalizace byl implementován ve skriptu *Optimalizace.m* následujícím způsobem:

- 1) *uživatelské nastavení optimalizace (viz Obrázek 6.3)*
- 2) *načtení a úprava dat*
- 3) *deklarace a inicializace proměnných (především všechny povolené kombinace parametrů a proměnná pro export dat)*
- 4) *proces optimalizace*
 - a. *for jednotlivé kombinace parametrů*
 - i. *for jednotlivé modely*
 1. *klouzavé průměry nebo jádrové odhady s aktuálními parametry pomocí funkce odhadni_funkci*
 2. *generování signálů pomocí funkce generuj_signaly*
 3. *eliminace opakovaných (zbytečných) signálů pomocí funkce vyfiltruj_signaly*
 4. *relativní zisky z každého obchodu a jejich medián*
 - ii. *end (for jednotlivé modely)*
 - iii. *maximální medián zisků a určení optimálního modelu pro danou kombinaci parametrů*
 - b. *end (for jednotlivé kombinace parametrů)*
- 5) *případný export do souboru Optimalizace-výstup.xlsx*
- 6) *optimální model a parametry pro zvolený akciový titul*

Funkce *odhadni_funkci.m* je definována takto⁸⁴:

$$[\text{odhad}] = \text{odhadni_funkci}(\text{rezim}, \text{data}, \text{x}, \text{h}, \text{druhOdhadu}, \text{zpozdeny}, \text{zpozdeni}) \quad (46)$$

kde *odhad* je jádrový odhad pro zadaná data – buď funkce hustoty (vzorec (3)) nebo regresní křivky (Nadaraya-Watsonův vzorec (22)), („double“; pole),
rezim je režim spuštění funkce – buď „hustota“ nebo „regrese“, („string“),
data jsou odhadovaná data, („double“; pole),
x jsou hodnoty, ve kterých má být odhad proveden, („double“; pole),
h je vyhlazovací parametr, („double“),

⁸⁴ V závorkách je u každého vstupu a výstupu funkce vždy uveden datový typ proměnné podle *MATLABu*.

druhOdhadu je jádrová funkce, která bude k odhadu použita⁸⁵ – pro potřeby skriptu buď „s“, „w“, „x“, „r“, „p“, „e“, „t“ nebo „g“, („string“),
 zpozdeny je proměnná určující, zda půjde o „zpožděný“ jádrový odhad (viz část 6.1.2.2), („logical“), a
 zpozdeni je parametr „posunu“, („double“).

V případě, že v (46) není specifikováno zpoždění je předpokládána rovnost vyhlazovacího parametru a parametru „posunu“. Funkce generuj_signaly.m je definována takto:

```
[nakup, prodej, nakupVarovani, prodejVarovani, zisky, poplatky, stav_uctu]=generuj_signaly(y1, y2, y3, h, rezim, stav_uctu, kurzy) (47)
```

kde **nakup** jsou vygenerované signály k nákupu podle pravidel (40), („logical“; pole),
prodej jsou vygenerované signály k prodeji podle pravidel (40), („logical“; pole),
nakupVarovani jsou vygenerované varovné signály k nákupu podle pravidel (40), („logical“; pole),
prodejVarovani jsou vygenerované varovné signály k prodeji podle pravidel (40), („logical“; pole),
zisky jsou zisky z jednotlivých uzavřených obchodů (výstup pouze v režimu „obchod“), („double“; pole),
poplatky jsou poplatky za realizaci jednotlivých signálů na trhu (výstup pouze v režimu „obchod“), („double“; pole),
stav_uctu je stav obchodního účtu po provedení všech vygenerovaných příkazů k nákupu a prodeji (výstup pouze v režimu „obchod“), („double“),
y1 je buď pole kurzů akcie nebo odhad krátkodobého charakteru, („double“; pole),
y2 je buď odhad krátkodobého (v případě, že y1 je kurz), střednědobého (v případě použití tří odhadů) nebo dlouhodobého charakteru (y1 je odhad krátkodobého charakteru), („double“; pole),
y3 je odhad dlouhodobého charakteru, („double“; pole),
h je parametr posunu y2 nebo y3 (podle použitých odhadů), („double“),
rezim je režim spuštění funkce – buď generování signálů („signaly“) nebo obchodní systém („obchod“), („string“),
stav_uctu je počáteční stav obchodního účtu („double“) a
kurzy je pole kurzů akcie („double“; pole).

Funkci (47) lze spustit v režimu generování signálů i obchodního systému, a to pro různý počet odhadů podle analýzy v části 4.3.1.5. Funkce (47) dále používá podpůrný skript *vypocitej_poplatek.m*, který vypočítá poplatek za realizaci pokynu na trhu podle Tabulka 2. Funkce *vyfiltruj_signaly.m* je definována takto:

```
[nakupDatумы, prodejDatумы]=vyfiltruj_signaly(nakupDatумы, prodejDatумы) (48)
```

kde **nakupDatумы** jsou datумы všech signálů k nákupu podle funkce *generuj_signaly.m* (47), následně vyfiltrované, („datetime“; pole) a
prodejDatумы jsou datумы všech signálů k prodeji podle funkce *generuj_signaly.m* (47), následně vyfiltrované, („datetime“; pole).

Funkce (48) vynechá opakované signály k nákupu do dalšího signálu k prodeji (ponechá 1. signál k nákupu) a vynechá opakované signály k prodeji, kdy není, co znovu prodat (ponechá 1. signál k prodeji).

⁸⁵ V případě klouzavých průměrů se vychází z angličtiny, v ostatních případech se zadává začáteční písmeno modelu, tj. jádrové funkce modelu.

6.3 Simulace obchodování

Po nalezení „vhodného“ modelu s „vhodným“ nastavením parametrů je již možné toto nastavení použít pro simulaci obchodování s jednotlivými akciovými tituly. Předtím je ale nezbytné určit strategii, podle které se nákup a prodej akcií bude řídit.

6.3.1 Předpoklady pro obchodování

Prvním důležitým předpokladem je, že obchodovány budou pouze dlouhé pozice. Krátké pozice nejsou v této práci uvažovány. Dále nechť je předpokládáno otevření obchodního účtu u obchodníka s disponibilní částkou 500 000 Kč. Pro každou realizaci pokynu jsou uvažovány poplatky uvedené v Tabulka 2.

Simulace probíhá pro každý akciový titul zvlášť, vždy v rámci stejného časového období a vždy se stejným počátečním vkladem. Žádný dodatečný vklad není povolen. Do každé dlouhé pozice, která je otevřena na doporučení obchodního systému, je investována maximální možná částka omezená pouze aktuálním stavem obchodního účtu, tj. nelze si půjčovat peníze od obchodníka ani nikoho jiného a nelze se dostat do záporného stavu. Objem každého obchodu je tedy, dle doporučení vycházejícího z analýzy poplatků v podkapitole 3.1, maximalizován.

Signály generované vytvořeným obchodním systémem zároveň reprezentují podané pokyny, tj. reakce na tyto signály je bezprostřední.

Výsledky získané pomocí vytvořených „optimálních“ modelů jsou porovnávány s „naivní“ strategií nazvanou „nakup-drž-prodej“. Použití této strategie znamená, že nejsou uvažovány žádné signály k nákupu nebo prodeji, ani není zohledňován vývoj kurzu akcií. Akcie jsou nakoupeny hned první burzovní den definovaného období (opět za maximální možnou částku) a jsou prodány poslední burzovní den tohoto období.

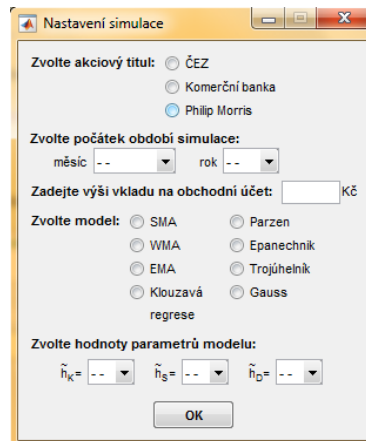
Pro srovnávání obou strategií může sloužit např. vyjádření celkového relativního zisku (resp. ztráty) pomocí vzorce

$$z^{[+/-]} = \frac{U + Z}{U} - 1, \quad (49)$$

kde $z^{[+/-]} \in \mathbb{R}$ je celkový relativní zisk (resp. ztráta) z obchodování po skončení celého období,
 $U \in \mathbb{R}^+$ je počáteční vklad na obchodní účet (počáteční stav) a
 $Z \in \mathbb{R}$ je celkový absolutní zisk (resp. ztráta) z obchodování po skončení celého období.

6.3.2 Popis algoritmu

K simulacím je použit příložený skript *Obchodni_system.m*. Po spuštění tohoto skriptu je zavolán pomocný skript *nastaveni_simulace.m*. Ten zobrazí menu, ve kterém uživatel zvolí akciový titul, pro který chce simulaci obchodování provést, dále volí model a jeho parametry. Poté zvolí počátek období, v rámci kterého chce simulaci provést (konec období je vždy poslední datum, pro který je kurz daného akciového titulu dostupný), a nakonec částku reprezentující počáteční stav obchodního účtu. Toto menu je zachyceno na Obrázek 6.6.



Obrázek 6.6: Menu pro nastavení obchodního systému (zdroj: vlastní implementace)

Obchodní systém byl implementován ve skriptu *Obchodni_system.m* následujícím způsobem:

- 1) *uživatelské nastavení simulace (viz Obrázek 6.6)*
- 2) *načtení a úprava dat*
- 3) *výpočet pro naivní strategii obchodování*
- 4) *tři klouzavé průměry nebo jádrové odhady s „optimálními“ parametry pomocí funkce odhadni_funkci (viz funkce (46))*
- 5) *vlastní obchodní systém pomocí funkce generuj_signaly (viz funkce (47) spuštěna v režimu „obchod“)*
- 6) *graf průběhu simulace obchodního systému*
- 7) *výsledky obchodování pomocí vytvořeného modelu*
 - a. *medián zisku (resp. ztráty)*
 - b. *průměrný zisk (resp. ztráta)*
 - c. *absolutní zisk (resp. ztráta)*
 - d. *počet realizovaných (uzavřených) obchodů*
 - e. *poplatky ze všech realizovaných pokynů*

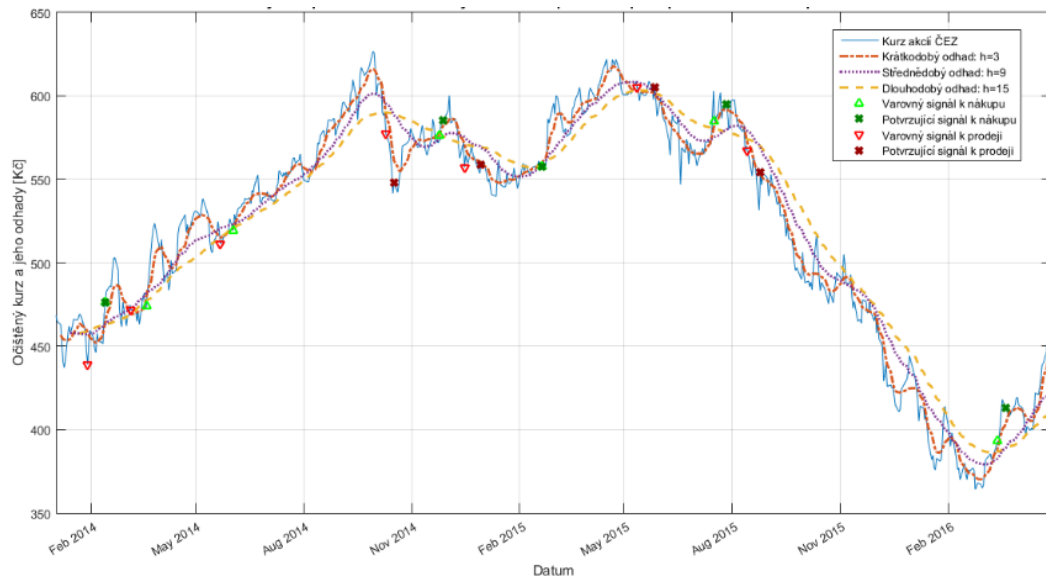
6.3.3 Akcie ČEZ

Obchodní systém provedl výpočty nejprve pro případ použití naivní strategie. Akcie byly nakoupeny 2. 1. 2014 a prodány 30. 4. 2016. Výsledek tohoto obchodu ukazuje Tabulka 7.

\tilde{c}_{N_1} [Kč]	\tilde{c}_{P_1} [Kč]	Počet uzavřených obchodů	Celkem za poplatky [Kč]	Z [Kč]	$z^{+/-}$
468.41,-	461.40,-	1	2 478.93,-	- 9 938.53,-	- 1.99 %

Tabulka 7: Výsledek naivního obchodu akcií ČEZ v rámci období od ledna 2014 do dubna 2016 (zdroj: vlastní simulace)

Aplikaci „optimální“ strategie pro obchodování akcií ČEZ znázorňuje Obrázek 6.7. Na obrázku je zachycen průběh obchodování pomocí modelu, který byl ze všech „optimálních“ nejvýnosnější (viz Tabulka 8). Vygenerované signály jsou zakresleny na křivce vývoje kurzu. Na období přelomu září a října 2014 je patrné, že, přestože je model nastaven tomuto akciovému titulu „na míru“, nedokáže identifikovat okamžik, kdy se kurz ocitá na svém maximu, a signál je vygenerován až, když se kurz propadl. To je však způsobeno konstrukcí odhadů kurzů tohoto typu, používaném v technické analýze, tj. zpožděné odhady. Zpoždění se může projevit negativně i tímto způsobem.



Obrázek 6.7: Obchodní systém pro akcie ČEZ založený na „optimálním“ modelu Epanechnik při kombinaci parametrů 3/9/15, $n = 582$ (zdroj: vlastní simulace)

Tabulka 8 uvádí výsledky obchodování při použití „optimálních“ strategií pro akcie ČEZ seřazené sestupně podle celkového relativního zisku $z^{+/-}$. Tečka (•) v tabulce značí libovolnou hodnotu z odpovídající množiny, ze které je parametr volen. Přestože obchodní systém neidentifikoval některé „nejlepší“ okamžiky pro otevření a uzavření dlouhé pozice (myšleno lokální maximum nebo minimum očištěného kurzu), jejich aproximace vedla k zisku více než 9 %. V porovnání s téměř 2% ztrátou při použití naivní strategie je na základě simulačního období výhodnější použít libovolný z optimalizovaných modelů.

„Optimální“ model	Kombinace parametrů	Počet uzavřených obchodů	Celkem za poplatky [Kč]	Medián $z_i^{+/-}$	Z [Kč]	$z^{+/-}$
Epanechnik	3/9/15	4	11 354.86,-	+ 1.51 %	45 215.32,-	+ 9.04 %
Epanechnik	3/8/16	4	11 313.41,-	+ 1.71 %	26 243.74,-	+ 5.25 %
Epanechnik	3/8/17	4	11 286.75,-	+ 0.82 %	16 587.38,-	+ 3.32 %
Gauss	3/12/13	4	11 271.73,-	+ 1.47 %	13 769.74,-	+ 2.75 %
Epanechnik	4/8/17	4	11 262.36,-	+ 0.81 %	9 530.88,-	+ 1.91 %
	5/8/17					
Gauss	3/11/14	4	11 258.11,-	+ 0.93 %	8 940.44,-	+ 1.79 %
Trojúhelník	•/9/16	4	11 241.91,-	- 0.15 %	2 720.95,-	+ 0.54 %
Gauss	•/10/16	4	11 225.90,-	- 0.15 %	- 2 492.35,-	- 0.50 %

Tabulka 8: Výsledek použití obchodního systému s „optimálním“ nastavením pro akcie ČEZ na období leden 2014 až duben 2016 (zdroj: vlastní simulace)

6.3.4 Akcie KB

Nejprve obchodní systém opět simuloval podání pokynů na základě výše uvedené naivní strategie, tj. nákup proběhl 2. 1. 2014 a prodej 29. 4. 2016. Tento uzavřený obchod shrnuje Tabulka 9.

\tilde{c}_{N_1} [Kč]	\tilde{c}_{P_1} [Kč]	Počet uzavřených obchodů	Celkem za poplatky [Kč]	Z [Kč]	$z^{+/-}$
3 842.96,-	4 859.00,-	1	2 492.26,-	128 577.34,-	+ 25.72 %

Tabulka 9: Výsledek naivního obchodu akcií KB v rámci období od ledna 2014 do dubna 2016 (zdroj: vlastní simulace)

Aplikace nejvýnosnější (viz Tabulka 10) z „optimálních“ strategií určených optimalizačním procesem v Tabulka 6 je znázorněna na Obrázek 6.8. Z obrázku je patrné, že systém, z důvodu krátkého časového rozmezí mezi střednědobým a dlouhodobým investičním horizontem, nereaguje na velké fluktuace kurzu dostatečně „pružně“. „Bezpečnost“ systému založená na varovných a potvrzujících signálech je v tomto případě velmi narušena. Až na první čtvrtletí roku 2015 nebyl v rámci simulačního období kurz „trendující“, ale spíše šlo o postranní trend, což může být, jak je naznačeno v částech 4.3.1.5 a 4.3.1.6, jednou z příčin vysokých ztrát po ukončení simulací (viz Tabulka 10).



Obrázek 6.8: Obchodní systém pro akcie KB založený na „optimálním“ modelu Trojúhelník při kombinaci parametrů 4/10/12, $n = 583$ (zdroj: vlastní simulace)

Tabulka 10 shrnuje výsledky obchodování při použití „optimálních“ strategií pro akcie KB seřazené sestupně podle celkového relativního zisku $z^{[+/-]}$. Použití „optimálních“ modelů v rámci simulačního období vedlo k více než 10% ztrátě. Naivní strategií bylo naopak dosaženo zisku téměř 26 %. Pro vyšší zisk by proto bylo zřejmě nutné upravit podmínky optimalizačního procesu.

„Optimální“ model	Kombinace parametrů	Počet uzavřených obchodů	Celkem za poplatky [Kč]	Medián $z_i^{[+/-]}$	Z [Kč]	$z^{[+/-]}$
Trojúhelník	4/10/12	8	20 960.23,-	- 1.21 %	- 53 401.88,-	- 10.68 %
	5/10/12					
Trojúhelník	3/10/12	7	18 420.46,-	- 1.80 %	- 69 752.96,-	- 13.95 %

Tabulka 10: Výsledek použití obchodního systému s „optimálním“ nastavením pro akcie KB na období leden 2014 až duben 2016 (zdroj: vlastní simulace)

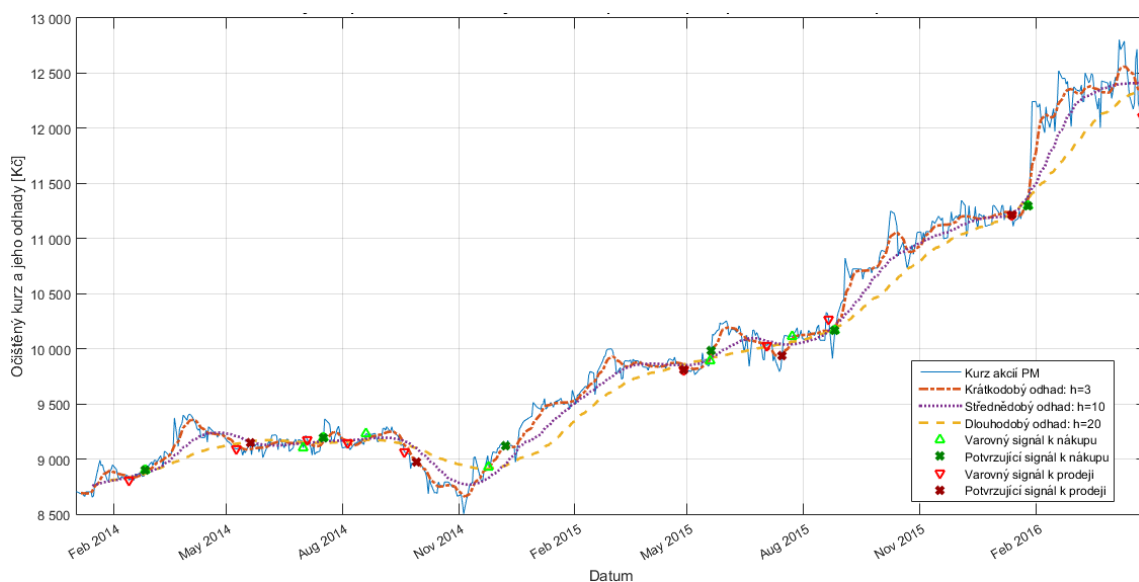
6.3.5 Akcie PM

Výsledky využití naivní strategie v případě obchodu s akciemi PM uvádí Tabulka 11. Nákup byl proveden 2. 1. 2014 a prodej 29. 4. 2016.

\tilde{c}_{N_1} [Kč]	\tilde{c}_{P_1} [Kč]	Počet uzavřených obchodů	Celkem za poplatky [Kč]	Z [Kč]	$z^{[+/-]}$
8 702.10,-	12 390.00,-	1	2 500.22,-	207 709.86,-	+ 41.54 %

Tabulka 11: Výsledek naivního obchodu akcií PM v rámci období od ledna 2014 do dubna 2016 (zdroj: vlastní simulace)

Průběh použití obchodního systému založeného na „optimální“ strategii obchodování s akciemi PM (viz Tabulka 6) je zachycen na Obrázek 6.9. V roce 2014 měl kurz spíše postranní charakter a nebyl nijak výrazně „trendující“, a proto jakákoliv větší fluktuace kurzu vedla k vygenerování signálu. Rostoucí trend se začal projevovat od přelomu let 2014 a 2015 a obchodní systém začal lépe fungovat. Pokud by nedošlo k prodeji akcií 13. 1. 2016., mohl by být zisk výrazně vyšší.



Obrázek 6.9: Obchodní systém pro akcie PM založený na „optimálním“ modelu Epanechnik při kombinaci parametrů 3/10/20, $n = 582$ (zdroj: vlastní simulace)

Výsledky obchodování s akciemi PM pomocí „optimálního“ modelu shrnuje Tabulka 12. Použití tohoto modelu vedlo k více než 15% zisku. Naivní strategie s více než 41% ziskem se však v rámci simulačního období ukázala být mnohem výnosnější.

„Optimální“ model	Kombinace parametrů	Počet uzavřených obchodů	Celkem za poplatky [Kč]	Medián $z_i^{[+/-]}$	Z [Kč]	$z^{[+/-]}$
Epanechnik	3/10/20	5	13 874.41,-	+ 2.23 %	76 682.16,-	+ 15.34 %

Tabulka 12: Výsledek použití obchodního systému s „optimálním“ nastavením pro akcie PM na období leden 2014 až duben 2016 (zdroj: vlastní simulace)

6.4 Modifikace optimalizačního procesu

Motivací k této kapitole je výsledek použití „optimálních“ modelů především v částech 6.3.4 a 6.3.5, kdy se jako výnosnější ukázala naivní strategie. V případě obchodování s akciemi KB se kromě postranního trendu ukázalo být problémem nastavení parametrů modelů, které vstupovaly do definovaného optimalizačního procesu. Přestože byly na základě definovaného testovacího období za nejvýhodnější vyhodnoceny kombinace parametrů uvedené v Tabulka 6, v rámci simulačního období toto nastavení nefungovalo tak dobře. Problémem může být rozdíl mezi jednotlivými investičními horizonty (v případě KB obzvláště mezi \tilde{h}_S a \tilde{h}_D , který byl pouze 2 obchodní dny).

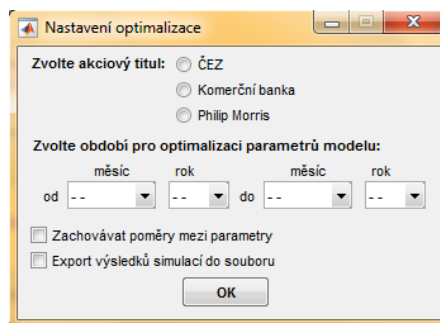
Pro optimalizační proces lze na základě výše popsané motivace upravit množiny \tilde{H}_K , \tilde{H}_S a \tilde{H}_D . V podkapitole 6.2 byly prvky jednotlivých množin určeny na základě poměrů mezi investičními horizonty doporučené kombinace parametrů 4/9/18, tj. 44,44 % a 50 %. Rozsah prvků množin byl podmíněn tím, aby nedošlo k jejich překrývání. Optimalizační proces prošel všechny kombinace parametrů bez dalších definovaných podmínek.

Výchozí množiny pro optimalizaci je proto možné upravit následujícím způsobem. Necht je doporučená kombinace 4/9/18 opět považována za výchozí, ale poměr mezi investičními horizonty nadále zachován. Proces volby „vhodné“ kombinace parametrů tedy bude probírat pouze několik vybraných možných kombinací, které splňují dané poměry mezi parametry. Kombinaci 4/9/18 lze

považovat za střed množiny uvažovaných kombinací. V tomto případě nelze uvažovat úpravu rozsahu na základě \tilde{h}_D , jelikož by mohly vycházet záporné \tilde{h}_K . Výchozí hodnota \tilde{h}_K je 4. Nechť je tedy minimální možná hodnota pro \tilde{h}_K rovna 2 (na základě výpočtu odhadů podle vzorců použitých pro jednotlivé modely). O odpovídající počet hodnot lze rozšířit množinu \tilde{H}_K také opačným směrem, tj. nechť je maximální možná hodnota pro \tilde{h}_K rovna 6. Určení hodnot zbývajících parametrů probíhá následujícím způsobem: $\tilde{h}_S = \frac{\tilde{h}_K}{0,4444} = \frac{2}{0,4444} \doteq 5$ a $\tilde{h}_D = \frac{\tilde{h}_S}{0,50} = \frac{5}{0,50} = 10$. Proces optimalizace tudíž bude procházet těchto pět kombinací parametrů zachovávajících poměry mezi parametry: 2/5/10, 3/7/14, 4/9/18, 5/11/22 a 6/14/28. Výsledné množiny jsou tedy následující: $\tilde{H}_K = \{2, 3, 4, 5, 6\}$, $\tilde{H}_S = \{5, 7, 9, 11, 14\}$ a $\tilde{H}_D = \{10, 14, 18, 22, 28\}$.

6.4.1 Optimalizace parametrů a modelů

Menu pro nastavení optimalizace z Obrázek 6.3 bylo pro potřeby modifikovaného optimalizačního procesu upraveno a lze v něm navíc volit mezi původním a modifikovaným způsobem. Nové menu je zachyceno na Obrázek 6.10.



Obrázek 6.10: Menu pro nastavení optimalizace upravené pro volbu zachování poměrů mezi parametry (zdroj: vlastní implementace)

Další postup je analogický optimalizačnímu principu popsanému v části 6.2.1. Algoritmus je proto také analogický s tím, který byl popsán v části 6.2.3. Upravena byla pouze část 3), ve které je na základě volby zachování poměrů odpovídajícím způsobem inicializována množina kombinací parametrů. Ostatní části algoritmu jsou pro oba přístupy shodné.

Tabulka 13 shrnuje výsledky optimalizace na testovacích datech při modifikované množině kombinací parametrů. Kompletní tabulky lze nalézt v souboru *Optimalizace-výstup.xlsx*. Při porovnání „optimálních“ modelů v Tabulka 6 a Tabulka 13 bylo zjištěno, že modifikace „vylepšila“ model pouze v případě ČEZ.

Akciový titul	Kombinace parametrů			Model	Medián $z_i^{[+/-]}$
	\tilde{h}_K	\tilde{h}_S	\tilde{h}_D		
ČEZ	6	14	28	Epanechnik	+ 3.02 %
KB	6	14	28	Klouzavá regrese	+ 0.98 %
PM	5	11	22	Epanechnik	+ 8.20 %

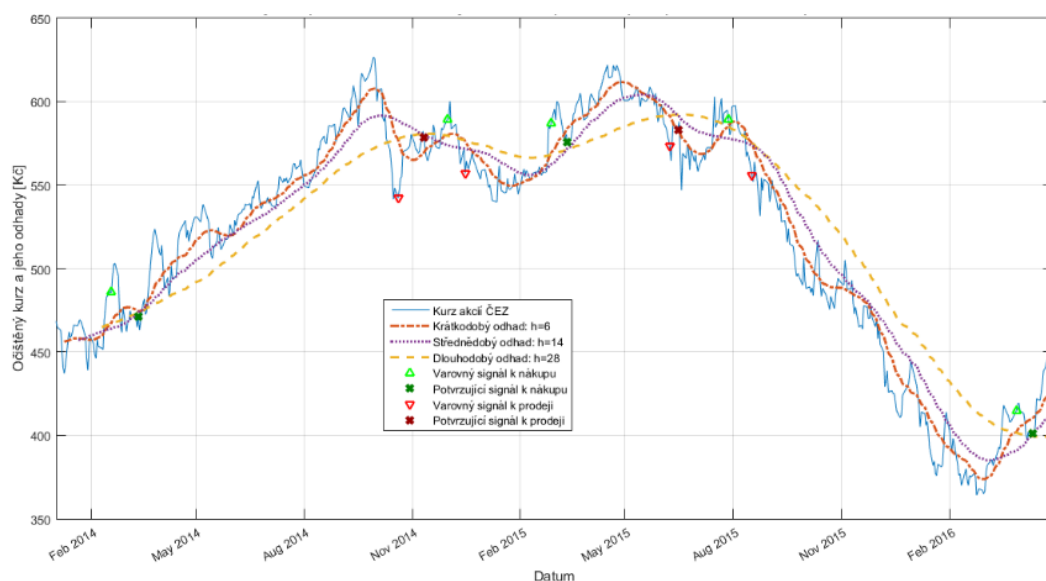
Tabulka 13: „Optimální“ obchodní modely pro jednotlivé akciové tituly při dodržování poměrů mezi parametry (zdroj: vlastní simulace)

6.4.2 Simulace obchodování

I v tomto případě jsou pro obchodování uvažovány předpoklady uvedené v části 6.3.1 a proto mohl také být využit algoritmus popsáný v části 6.3.2.

6.4.2.1 Akcie ČEZ

Průběh obchodování s akciemi ČEZ pomocí obchodního systému založeného na optimálním nastavení podle Tabulka 13 ukazuje Obrázek 6.11. Díky novému nastavení parametrů systém nevygeneroval signály pro obchody, které by v rámci simulačního období byly ztrátové, tj. obchod na přelomu let 2014 a 2015 a v srpnu 2015 (viz Obrázek 6.7). Zároveň díky prodloužení investičních horizontů nedošlo k vygenerování signálů v nevhodné okamžiky, např. během propadu kurzu na přelomu září a října 2014. S tím však také souvisí snížení zisku z druhého (druhý na Obrázek 6.11, třetí na Obrázek 6.7) obchodu.



Obrázek 6.11: Obchodní systém pro akcie ČEZ založený na „optimálním“ modelu Epanechnik při kombinaci parametrů 6/14/28, $n = 582$ (zdroj: vlastní simulace)

V Tabulka 14 je uveden výsledek simulace po modifikovaném výběru optimálního nastavení modelu a jeho parametrů. Díky snížení počtu uzavřených obchodů došlo k odpovídajícímu snížení poplatků za provedené pokyny. Jelikož oba uzavřené obchody byly ziskové, vzrostl také celkový relativní zisk za celé období, a to z 9 % na více než 23 %.

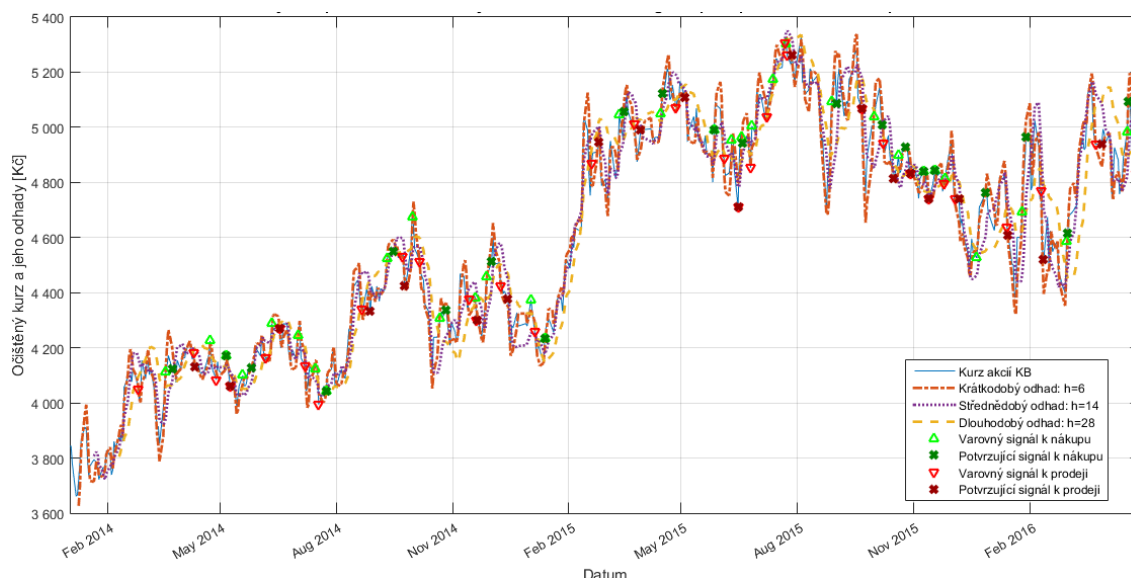
„Optimální“ model	Kombinace parametrů	Počet uzavřených obchodů	Celkem za poplatky [Kč]	Medián $z_i^{[+/-]}$	Z [Kč]	$z^{[+/-]}$
Epanechnik	6/14/28	2	6 533.30,-	+ 11.68 %	116 815.44,-	+ 23.36 %

Tabulka 14: Výsledek použití obchodního systému s „optimálním“ (na základě modifikovaného optimalizačního procesu) nastavením pro akcie ČEZ na období leden 2014 až duben 2016 (zdroj: vlastní simulace)

6.4.2.2 Akcie KB

Obrázek 6.12 zachycuje průběh použití obchodního systému pro akcie KB založeného na „optimálním“ nastavení podle Tabulka 13. Optimalizačním procesem byl vybrán model Klouzavá regrese. Při porovnání s použitím modelu Trojúhelník, který byl „optimální“ podle původní optimalizace, je z Obrázek 6.12 (a následně také z Tabulka 15) zřejmé, že došlo k uzavření mnohem většího počtu obchodů. S tím také souvisí zvýšení nákladů na poplatky v odpovídající výši. Jak je ale

vidět při srovnání posledního sloupce Tabulka 10 a Tabulka 15, celkový relativní zisk (v tomto případě ztráta) za celé simulační období vzrostl (resp. klesla) o jeden procentní bod, tj. o více než 4 000 Kč. Modifikací optimalizačního procesu tedy bylo dosaženo mírného zlepšení. V rámci simulačního období se však ukázala být úspěšnější definovaná naivní strategie.



Obrázek 6.12: Obchodní systém pro akcie KB založený na „optimálním“ modelu Klouzavá regrese při kombinaci parametrů 6/14/28, $n = 583$ (zdroj: vlastní simulace)

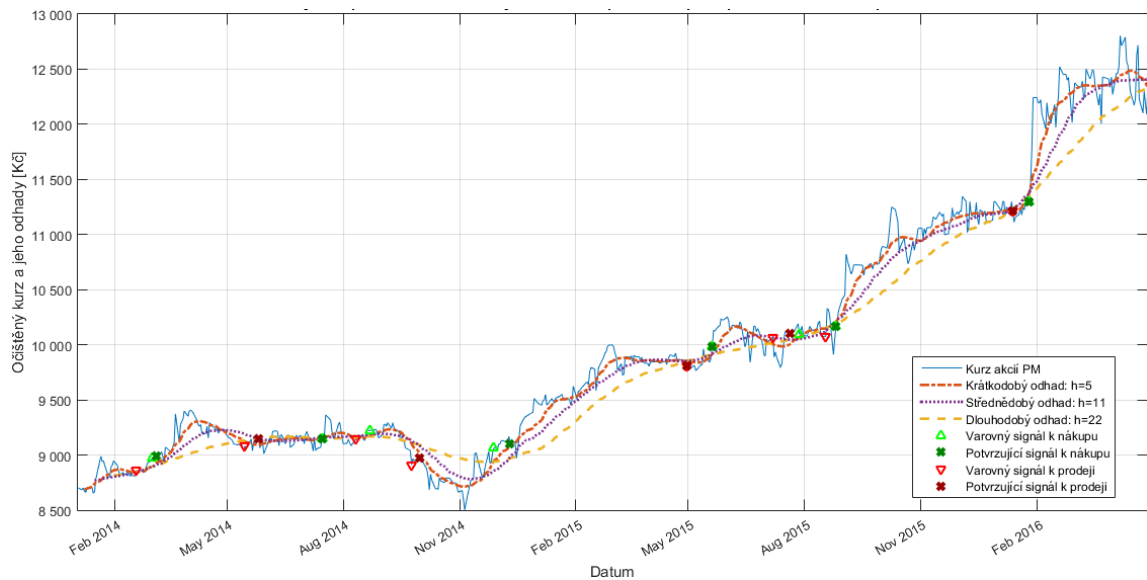
„Optimální“ model	Kombinace parametrů	Počet uzavřených obchodů	Celkem za poplatky [Kč]	Medián $z_i^{[+/-]}$	Z [Kč]	$z^{[+/-]}$
Klouzavá regrese	6/14/28	20	50 894.72,-	- 2.23 %	-49 193.92,-	- 9.84 %

Tabulka 15: Výsledek použití obchodního systému s „optimálním“ (na základě modifikovaného optimalizačního procesu) nastavením pro akcie KB na období leden 2014 až duben 2016 (zdroj: vlastní simulace)

6.4.2.3 Akcie PM

Použití obchodního systému pro akcie PM založeného na optimálním nastavení podle Tabulka 13 je zachyceno na Obrázek 6.13. Jelikož se u použitého modelu změnila pouze kombinace parametrů, je v tomto případě možné sledovat vliv změny těchto parametrů. Každý z investičních horizontů je mírně delší než u modelu použitého v části 6.3.5. Obchody byly uzavřeny přibližně ve stejných obdobích jako v případě Obrázek 6.9. Nové nastavení tedy upřesnulo okamžiky otevření a uzavření dlouhé pozice.

Tabulka 16 shrnuje použití daného obchodního systému pro akcie PM. Přestože, v porovnání s Tabulka 12, bylo s novým nastavením dosaženo nižšího mediánu zisku, celkový zisk za celé období vzrostl o více než 1.5 procentního bodu, tj. o více než 8 000 Kč. I v tomto případě se však v rámci simulačního období ukázala být úspěšnější definovaná naivní strategie.



Obrázek 6.13: Obchodní systém pro akcie PM založený na „optimálním“ modelu Epanechnik při kombinaci parametrů 5/11/22, $n = 582$ (zdroj: vlastní simulace)

„Optimální“ model	Kombinace parametrů	Počet uzavřených obchodů	Celkem za poplatky [Kč]	Medián $Z_i^{[+/-]}$	Z [Kč]	$Z^{[+/-]}$
Epanechnik	5/11/22	5	13 891.18,-	+ 1.20 %	84 865.61,-	+ 16.97 %

Tabulka 16: Výsledek použití obchodního systému s „optimálním“ (na základě modifikovaného optimalizačního procesu) nastavením pro akcie PM na období leden 2014 až duben 2016 (zdroj: vlastní simulace)

7 Závěr

Diplomová práce pojednává o problematice jádrových odhadů, obchodování s akciemi na burze cenných papírů a propojení obou těchto oblastí. Práci lze rozdělit do čtyř částí.

První část práce je souhrnem teoretických poznatků o neparametrických jádrových odhadech. Jsou uvedeny tři druhy odhadů podle jejich účelu, tj. odhad funkce hustoty, distribuční funkce a regresní křivky. V každé této části je uvedena motivace použití jádrových odhadů, která má vždy stejnou strukturu, kdy je nejprve demonstrováno použití parametrických, dále jednoduchých neparametrických a nakonec jádrových odhadů. Každá fáze je, v porovnání s předchozí, považována za „vylepšení“ přístupu k odhadování. Pozornost je také věnována problematice volby vyhlazovacího parametru u jádrových odhadů. Z důvodu zaměření práce je v závěru kapitoly odvozen dále používaný Nadaraya-Watsonův jádrový odhad regresní křivky.

Druhá část popisuje zvolenou aplikační oblast, tj. obchodování s akciemi na Burze cenných papírů Praha. Nejprve je seznámeno se způsobem obchodování s akciemi. Jsou popsány náklady na obchodování, které jsou tvořeny především poplatky za realizaci pokynů k nákupu a prodeji akcií. Dále v případě zisku z uzavřeného obchodu a splnění zákonem definovaných podmínek může být za náklad považována také daň z příjmu.

Hlavní částí aplikační oblasti však je disciplína zvaná technická analýza. Pro potřeby této práce je pozornost věnována především technickým indikátorům, konkrétně různým druhům klouzavých průměrů. Motivací k použití jádrových odhadů v této oblasti je způsob použití několika klouzavých průměrů, jejichž překřížení signalizuje okamžiky „vhodné“ k nákupu nebo prodeji akcií.

Ve třetí části práce jsou popsány vybrané akciové tituly, na kterých je dále testováno použití navrhovaného obchodního systému. Je zde popsán způsob úpravy kurzů (získaných z veřejně a bezplatně dostupných zdrojů), které je třeba před použitím k další analýze očistit od možného štěpení akcií a výplat dividendy. Každý akciový titul je stručně popsán a zachycen vývoj kurzu očištěného od výše popsaných vlivů.

Ve čtvrté a poslední části jsou popsány navržené modely a způsoby odhadu jejich parametrů. Místo metody křížového ověřování popsané ve druhé kapitole je k odhadu vyhlazovacího parametru u jádrového odhadu navržen odhad rovnající se parametru posunu. Použití této metodiky je následně úspěšně ověřeno. Jádrové odhady jsou díky tomuto přístupu použitelné analogickým způsobem jako klouzavé průměry v technické analýze a jejich způsob použití pro obchodní systém je proto shodný. Dále je navržen obchodní systém založený na třech křivkách stejného typu lišících se pouze hodnotou parametrů posunu (resp. zpoždění), které byly zvoleny na základě přístupu technické analýzy k obchodování s akciemi.

Praktická část práce je rozdělena do dvou částí. Nejprve je popsán a proveden numerický optimalizační proces v rámci testovacího období, jehož výstupem jsou konkrétní modely s kombinacemi parametrů „vhodné“ pro obchodování s akciemi dané akciové společnosti, tj. modely a parametry jsou určeny „na míru“ akciovému titulu. Během optimalizace bylo mj. ukázáno, že stejně jako volba parametrů jádrových odhadů, také volba jádrové funkce může být

významná. Tyto modely jsou následně použity jako nastavení obchodního systému, který je testován na datech od roku 2014 do konce dubna 2016. Na základě získaných výsledků je následně upravena množina kombinací parametrů, které vstupují do procesu optimalizace, a celý postup je proveden znovu. Touto modifikací bylo dosaženo lepších výsledků (tj. vyšších zisků (resp. nižších ztrát)) při použití „optimálních“ modelů.

Přestože se v případě dvou akciových titulů projevila jako úspěšnější naivní strategie „nakup-drž-prodej“, je třeba brát v úvahu, že tato strategie nezohledňuje jakékoliv výkyvy kurzu. Jejím použitím se tak investor vystavuje velkému riziku ztráty. Navržený obchodní systém naopak reaguje na pohyb kurzu akcií, a proto ho lze považovat za mnohem „bezpečnější“ (nejen na základě použití systému varovný-potvrzující signál). Simulace navíc ukázaly, že použití takového systému může vést k nemalým ziskům. Na druhou stranu, jeho použití pro akcie KB lze považovat za případ varující před „slepým“ používáním takto definovaného systému. Pro dosažení lepších výsledků i v tomto případě by např. bylo možné optimalizovat parametry po každém uzavřeném obchodu, tj. po realizaci pokynu k prodeji akcií a následně každý obchodní den provádět simulace obchodování a čekat na pokyn k prodeji. Mohlo by tedy být voleno pružné období optimalizace, jehož počátek by mohl být omezen pouze dostupností historických kurzů akcií. Dále by obchodní systém mohl být vylepšen např. dalšími technickými indikátory, kdy by srovnával a vyhodnocoval pokyny vygenerované na základě několika indikátorů. Mohla by také přibýt možnost zadání „chytrých“ pokynů, které mohou buď reagovat na výši aktuální ztráty nebo i zisku.

Při vytváření obchodního systému je nutné si uvědomit, že univerzální a 100% funkční systém není možné vytvořit. Kurz každého akciového titulu se pohybuje odlišným způsobem a jeho předpověď je komplikovaná záležitost. Každý investor je navíc ochotný podstupovat různé míry rizika. Nejen na základě toho je tvorba obchodního systému velice individuální otázkou. V této práci byl testován alternativní přístup k obchodování s akciemi podle technické analýzy, který se ukázal jako úspěšnější a použitelný pro reálné obchodování.

Seznam použitých zkratek a značení

n	... počet hodnot (pozorování); vždy je předpokládáno $n \in \mathbb{N}$
\mathbb{N}	... množina přirozených čísel, tj. 1, 2, 3, 4, ...
$f(x)$... funkce hustoty náhodné veličiny X
$\hat{f}(x)$... odhad funkce hustoty sestrojený pomocí histogramu
\mathbb{R}^+	... množina kladných reálných čísel
\mathbb{R}	... množina reálných čísel
$N(0,1)$... normální rozdělení pravděpodobnosti se střední hodnotou 0 a rozptylem 1
x_{max}	... $\max_x x_i$
x_{min}	... $\min_x x_i$
s_X	... výběrová směrodatná odchylka pro náhodnou veličinu X
IQR	... kvartilová odchylka
$x_{0.75}$... 75% kvantil náhodné veličiny X
$k(x)$... jádrová funkce
i.i.d. náhodné veličiny	... nezávislé, stejně rozdělené náhodné veličiny (z angl. "independent and identically distributed")
$\hat{f}_h(x)$... jádrový odhad funkce hustoty při hodnotě vyhlazovacího parametru h
$E(X)$... střední hodnota náhodné veličiny X
MSE	... střední kvadratická chyba (z angl. Mean Square Error)
MISE	... střední integrální kvadratická chyba (z angl. Mean Integrated Square Error)
AMISE	... asymptotická střední integrální kvadratická chyba (z angl. Asymptotic Mean Integrated Square Error)
h_{opt}	... teoretická optimální hodnota vyhlazovacího parametru h
$x_{(i)}$... i -té pozorování ze seřazeného výběru náhodné veličiny X
H_n	... množina zahrnující všechny uvažované hodnoty vyhlazovacího parametru pro odhad h_{opt}
\hat{h}	... odhad h_{opt} pomocí metody křížového ověřování
$CV_f(h)$... funkce křížového ověřování pro funkci hustoty f
$\hat{f}_{h,-i}(x_i)$... jádrový odhad funkce hustoty v bodě x_i bez použití i -tého pozorování
$F(x)$... distribuční funkce náhodné veličiny X
EDF	... empirická distribuční funkce
$K(x)$... jádrová distribuční funkce
$m(x)$... regresní křivka
ε	... nesystematická složka regresního modelu
$D(X)$... rozptyl náhodné veličiny X
$C(X, Y)$... kovariance náhodných veličin X a Y
$\hat{m}_{SMA}(x_i)$... odhad regresní křivky pomocí jednoduchého klouzavého průměru (z angl. Simple Moving Average)
$A_1 \cap A_2$... průnik množin A_1 a A_2
\emptyset	... prázdná množina
$f(x, y)$... sdružená hustota pravděpodobnosti náhodných veličin X a Y
$E(X Y)$... podmíněná střední hodnota
e_i	... reziduum při odhadu j -tého pozorování vysvětlované proměnné
$f_X(x)$... marginální (okrajová) hustota pravděpodobnosti náhodné veličiny X
$\hat{f}_X(x)$... odhad funkce $f_X(x)$

$\xrightarrow{n \rightarrow \infty}$... $\lim_{n \rightarrow \infty}$
\rightarrow	... zkrácený zápis pro konvergenci výrazu (pro potřeby důkazů)
$D(f)$... je definiční obor funkce f
$\hat{f}(x, y)$... odhad funkce $f(x, y)$
$k_*(x, y)$... dvourozměrná jádrová funkce
$\hat{m}_h(x)$... Nadaraya-Watsonův jádrový odhad regresní funkce při hodnotě vyhlazovacího parametru h
κ	... řád jádrové funkce
AMSE	... průměrná střední kvadratická chyba (z angl. Average Mean Square Error)
RSS	... reziduální součet čtverců (z angl. Residual Sum of Squares)
\mathbb{N}_0	... množina přirozených čísel a nula, tj. 0, 1, 2, 3, 4, ...
$CV_m(h)$... funkce křížového ověřování pro regresní funkci m
$\hat{m}_{h,-i}(x_i)$... Nadaraya-Watsonův jádrový odhad regresní funkce v bodě x_i (3) bez použití i -tého pozorování
BCPP	... Burza cenných papírů Praha, a.s.
EU	... Evropská Unie
Fio	... Fio banka a.s.
$S_h(t)$... jednoduchý klouzavý průměr délky h
c_t	... kurz akcií k datu t
ČEZ	... ČEZ a.s.
$W_h(t)$... vážený klouzavý průměr délky h
$E_h(t)$... exponenciální klouzavý průměr délky h
\hat{c}_t	... Odhad kurzu akcií k datu t pomocí klouzavé regrese
DK_{t_i}	... dividendový koeficient založený na dividendě s ex-dividendovým datem t_i
d_{t_i}	... výše dividendy k datu t_i
\tilde{c}_t	... očištěný kurz akcií od případné dividendy a štěpení akcií k datu t
SK_s	... štěpící koeficient založený na štěpení akcií k datu s
O2	... O2 Czech Republic, a.s.
KB	... Komerční banka, a.s.
SMA	... jednoduchý klouzavý průměr (z angl. Simple Moving Average)
WMA	... vážený klouzavý průměr (z angl. Weighted Moving Average)
EMA	... exponenciální klouzavý průměr (z angl. Exponential Moving Average)
\tilde{h}_K	... parametr posunu modelu pro krátkodobý investiční horizont
\tilde{h}_S	... parametr posunu modelu pro střednědobý investiční horizont
\tilde{h}_D	... parametr posunu modelu pro dlouhodobý investiční horizont
$\tilde{h}_K / \tilde{h}_S / \tilde{h}_D$... zápis parametrů modelu pro jednotlivé investiční horizonty
$z_i^{[+/-]}$... relativní zisk (resp. ztráta) z i -tého uzavřeného obchodu
\tilde{c}_{P_i}	... kurz, za který jsou akcie v i -tém uzavřeném obchodě prodány
\tilde{c}_{N_i}	... kurz, za který jsou akcie v i -tém uzavřeném obchodě nakoupeny
$z^{[+/-]}$... celkový relativní zisk (resp. ztráta) z obchodování po skončení simulace obchodního systému
U	... počáteční vklad na obchodní účet (počáteční stav)
Z	... celkový absolutní zisk (resp. ztráta) z obchodování po skončení simulace obchodního systému

Seznam obrázků

Obrázek 2.1: Příklad histogramu pro hodnocení slanosti 16% krémoveho sýru, pro výpočet počtu tříd použito Sturgesovo pravidlo, $h = 0.5250$, $n = 24$ (zdroj: vlastní zpracování).....	4
Obrázek 2.2: Srovnání histogramů s různými počty (resp. šířkami – vepsané v legendách u obrázků) tříd a známé hustoty, $n = 1\ 000$ (zdroj: vlastní simulace).....	4
Obrázek 2.3 Graficky zachycené průběhy vybraných jádrových funkcí kx (zdroj: vlastní simulace)	6
Obrázek 2.4: Příklad konstrukce jádrového odhadu hustoty za použití Epanechnikovy jádrové funkce, $h = 0.9$, $n = 7$ (zdroj: vlastní simulace).....	7
Obrázek 2.5: Průběh funkce křížového ověřování za použití Epanechnikovy jádrové funkce s vyznačenou minimální hodnotou $h = 0.3530$, $n = 1000$, krok simulace nastaven na 0.0001 (zdroj: vlastní simulace).....	9
Obrázek 2.6: Znamá hustota a její odhady v podobě histogramu a jádrového odhadu za použití Epanechnikovy jádrové funkce a metody křížového odhadu vyhlazovacího parametru, $h = 0.3530$, $n = 1\ 000$ (zdroj: vlastní simulace).....	10
Obrázek 2.7: Nevhodná volba regresního modelu – srovnání odhadu a teoretické regresní křivky, $n = 26$ (zdroj: vlastní simulace).....	12
Obrázek 2.8: Odhad regresní křivky pomocí regresogramu s ekvidistantními intervaly, $h = 0.4167$, $n = 26$ (zdroj: vlastní simulace).....	13
Obrázek 2.9: Odhad regresní křivky pomocí tří jednoduchých klouzavých průměrů různých délek, $h_1 = 3$, $h_2 = 7$, $h_3 = 11$, $n = 26$ (zdroj: vlastní simulace).....	15
Obrázek 2.10: Nadaraya-Watsonův jádrový odhad regresní křivky za použití Epanechnikovy jádrové funkce a metody křížového odhadu vyhlazovacího parametru, $h = 0.2191$, $n = 26$ (zdroj: vlastní simulace).....	20
Obrázek 2.11: Průběh funkce křížového ověřování za použití Epanechnikovy jádrové funkce s vyznačenou minimální hodnotou $h = 0.2191$, $n = 26$, krok simulace nastaven na 0.0001 (zdroj: vlastní simulace).....	23
Obrázek 3.1: Vývoj relativní výše poplatků při obchodování českých akciových titulů na BCPP prostřednictvím Fio (zdroj: vlastní simulace).....	25
Obrázek 3.2: Vývoj absolutní výše poplatků při obchodování českých akciových titulů na BCPP prostřednictvím Fio (zdroj: vlastní simulace).....	25
Obrázek 4.1: Jednoduché klouzavé průměry různých délek klouzavých oken spočítané z kurzu akcií ČEZ očištěného o dividendy, $n = 833$ (zdroj: vlastní simulace).....	30
Obrázek 4.2: Váhy přiřazované jednotlivým kurzům akcií v čase při použití SMA, WMA a EMA, $h = 10$, $n = 13$ (zdroj: vlastní simulace).....	31
Obrázek 4.3: Generování signálů k nákupu/prodeji akcií ČEZ pomocí jednoduchého klouzavého průměru, $h = 4$, $n = 83$ (zdroj: vlastní simulace).....	34
Obrázek 4.4: Generování signálů k nákupu/prodeji akcií ČEZ pomocí krátkodobého ($h = 4$) a dlouhodobého ($h = 18$) jednoduchého klouzavého průměru, $n = 269$ (zdroj: vlastní simulace).....	35
Obrázek 4.5: Generování signálů k nákupu/prodeji akcií ČEZ pomocí krátkodobého ($h = 4$), střednědobého ($h = 9$) a dlouhodobého ($h = 18$) jednoduchého klouzavého průměru, $n = 269$ (zdroj: vlastní simulace).....	36

Obrázek 5.1: Vývoj kurzu akcií ČEZ očištěného o dividendy, $n = 3348$ (zdroj dat: [44], záložka „Vývoj cen akcií“, a [45])	39
Obrázek 5.2: Vývoj kurzu akcií KB očištěného o dividendy, $n = 3663$ (zdroj dat: [47], záložka „Akcie KB“)	40
Obrázek 5.3: Vývoj kurzu akcií PM očištěného o dividendy, $n = 5521$ (zdroj dat: [49])	41
Obrázek 6.1: Obchodní systém pro akcie ČEZ používající jádrový odhad s Epanechnikovou jádrovou funkcí, odhad parametru h metodou křížového ověřování, $h = 0.7256$, $n = 83$ (zdroj: vlastní simulace)	43
Obrázek 6.2: Obchodní systém pro akcie ČEZ používající jádrový odhad s Epanechnikovou jádrovou funkcí (krátkodobý horizont), odhad parametru h metodou křížového ověřování, $h = 4$, $n = 83$ (zdroj: vlastní simulace).....	44
Obrázek 6.3: Menu pro nastavení optimalizace (zdroj: vlastní implementace)	46
Obrázek 6.4: Obchodní systém pro akcie ČEZ aplikovaný na testovací data od ledna 2009 do prosince 2013 používající exponenciální klouzavé průměry, kombinace parametrů 4/9/18 (zdroj: vlastní simulace)	48
Obrázek 6.5: Obchodní systém pro akcie ČEZ aplikovaný na testovací data od ledna 2009 do prosince 2013 používající jádrový odhad s Gaussovou jádrovou funkcí, kombinace parametrů 4/9/18 (zdroj: vlastní simulace)	49
Obrázek 6.6: Menu pro nastavení obchodního systému (zdroj: vlastní implementace).....	53
Obrázek 6.7: Obchodní systém pro akcie ČEZ založený na „optimálním“ modelu Epanechnik při kombinaci parametrů 3/9/15, $n = 582$ (zdroj: vlastní simulace)	54
Obrázek 6.8: Obchodní systém pro akcie KB založený na „optimálním“ modelu Trojúhelník při kombinaci parametrů 4/10/12, $n = 583$ (zdroj: vlastní simulace).....	55
Obrázek 6.9: Obchodní systém pro akcie PM založený na „optimálním“ modelu Epanechnik při kombinaci parametrů 3/10/20, $n = 582$ (zdroj: vlastní simulace).....	56
Obrázek 6.10: Menu pro nastavení optimalizace upravené pro volbu zachování poměrů mezi parametry (zdroj: vlastní implementace)	57
Obrázek 6.11: Obchodní systém pro akcie ČEZ založený na „optimálním“ modelu Epanechnik při kombinaci parametrů 6/14/28, $n = 582$ (zdroj: vlastní simulace).....	58
Obrázek 6.12: Obchodní systém pro akcie KB založený na „optimálním“ modelu Klouzavá regrese při kombinaci parametrů 6/14/28, $n = 583$ (zdroj: vlastní simulace).....	59
Obrázek 6.13: Obchodní systém pro akcie PM založený na „optimálním“ modelu Epanechnik při kombinaci parametrů 5/11/22, $n = 582$ (zdroj: vlastní simulace)	60

Seznam tabulek

Tabulka 1: Vybrané jádrové funkce hustoty (pro výpočty viz [10], str. 38-39).....	6
Tabulka 2: Přehled poplatků za uzavřené obchody s českými akciovými tituly při obchodování na BCPP prostřednictvím Fio (zdroje: [26], str. 5, [28] a [29])	25
Tabulka 3: Data pro první čtyři doporučené obchody s akciemi ČEZ od ledna 2009 do prosince 2013, srovnání jednotlivých modelů s parametry 4/9/18 (zdroj: vlastní simulace).....	47
Tabulka 4: Výsledky doporučeného obchodování s akciemi ČEZ v rámci testovacího období, srovnání jednotlivých modelů s parametry 4/9/18 (zdroj: vlastní simulace)	48
Tabulka 5: Příklad vložení záznamu o optimálním modelu pro danou kombinaci parametrů do tabulky (zdroj: vlastní soubor <i>Optimalizace-výstup.xlsx</i>).....	49
Tabulka 6: „Optimální“ obchodní modely pro jednotlivé akciové tituly (zdroj: vlastní simulace)...	50
Tabulka 7: Výsledek naivního obchodu akcií ČEZ v rámci období od ledna 2014 do dubna 2016 (zdroj: vlastní simulace).....	53
Tabulka 8: Výsledek použití obchodního systému s „optimálním“ nastavením pro akcie ČEZ na období leden 2014 až duben 2016 (zdroj: vlastní simulace).....	54
Tabulka 9: Výsledek naivního obchodu akcií KB v rámci období od ledna 2014 do dubna 2016 (zdroj: vlastní simulace).....	54
Tabulka 10: Výsledek použití obchodního systému s „optimálním“ nastavením pro akcie KB na období leden 2014 až duben 2016 (zdroj: vlastní simulace).....	55
Tabulka 11: Výsledek naivního obchodu akcií PM v rámci období od ledna 2014 do dubna 2016 (zdroj: vlastní simulace)	55
Tabulka 12: Výsledek použití obchodního systému s „optimálním“ nastavením pro akcie PM na období leden 2014 až duben 2016 (zdroj: vlastní simulace).....	56
Tabulka 13: „Optimální“ obchodní modely pro jednotlivé akciové tituly při dodržování poměrů mezi parametry (zdroj: vlastní simulace)	57
Tabulka 14: Výsledek použití obchodního systému s „optimálním“ (na základě modifikovaného optimalizačního procesu) nastavením pro akcie ČEZ na období leden 2014 až duben 2016 (zdroj: vlastní simulace)	58
Tabulka 15: Výsledek použití obchodního systému s „optimálním“ (na základě modifikovaného optimalizačního procesu) nastavením pro akcie KB na období leden 2014 až duben 2016 (zdroj: vlastní simulace).....	59
Tabulka 16: Výsledek použití obchodního systému s „optimálním“ (na základě modifikovaného optimalizačního procesu) nastavením pro akcie PM na období leden 2014 až duben 2016 (zdroj: vlastní simulace)	60

Reference

- [1] ANTOCH, J. a D. VORLÍČKOVÁ. *Vybrané metody statistické analýzy dat*. Praha: Academia, 1992, 280 s.. ISBN 80-200-0204-9.
- [2] HOROVÁ, I. Teorie a praxe jádrového vyhlazování. In: HOLČÍK, J. *Matematická biologie: e-learningová učebnice* [online]. KOMENDA, M. ed. Brno: Masarykova univerzita, 2015, s. 73 [cit. 2016-02-21]. ISBN 978-80-210-8095-9. Dostupné z: http://is.muni.cz/www/98951/41610771/43823411/43823458/Analyza_a_hodnoc/45606646/Teorie_a_praxe_jadroveho_vyhlazovani.pdf
- [3] HÄRDLE, W. et al. *Nonparametric and Semiparametric Models*. New York: Springer, 2004, 299 s.. ISBN 3-540-20722-8.
- [4] KOLÁČEK, J. *Jádrové odhady regresní funkce*. Brno: 2004. Disertační práce. Masarykova univerzita, Přírodovědecká fakulta. 84 s..
- [5] PAGAN, A. a A. ULLAH. *Nonparametric econometrics*. New York: Cambridge University Press, 1999, 424 s.. ISBN 0 521 58611 9.
- [6] NEUBAUER, J. M. SEDLAČÍK a O. KŘÍŽ. *Základy statistiky: Aplikace v technických a ekonomických oborech*. Praha: Grada, 2012, 240 s.. ISBN 978-80-247-4273-1.
- [7] SCOTT, D. W. ResearchGate. *Wiley Interdisciplinary Reviews: Computational Statistics* [online]. Wiley, 25. června 2010, **2** (4), 497-502 [cit. 2016-02-21]. ISSN 19395108. DOI: 10.1002/wics.103. Dostupné z: https://www.researchgate.net/publication/250594515_Scott's_rule
- [8] SCOTT, D. W. *Multivariate Density Estimation: Theory, Practice, and Visualization*. 2nd ed. Hoboken (New Jersey): Wiley, 2015, 384 s.. ISBN 978-0-471-69755-8.
- [9] TURLACH, B. A. Bandwidth Selection in Kernel Density Estimation: A Review. In: *ResearchGate* [online]. Louvain-la-Neuve (Belgium): 3. 8. 2013 [cit. 2016-02-21]. Dostupné z: https://www.researchgate.net/publication/2316108_Bandwidth_Selection_in_Kernel_Density_Estimation_A_Review
- [10] ŤOUPAL, T. *Neparametrický odhad spolehlivosti a odhad trendové složky*. Plzeň: 2013. Disertační práce. Západočeská univerzita v Plzni, Fakulta aplikovaných věd, Katedra matematiky. Vedoucí práce František Vávra. 138 s..
- [11] WASSERMAN, L. *All of Nonparametric Statistics*. Springer texts in statistics. New York: Springer, 2006, 268 s.. ISBN 0-387-25145-6.

- [12] ŘEZÁČ, M. *Jádrové odhady hustoty*. Brno: 2007. Disertační práce. Masarykova univerzita, Přírodovědecká fakulta. 75 s..
- [13] NOVÁKOVÁ, J. *Jádrové odhady distribuční funkce*. Brno: 2009. Diplomová práce. Masarykova univerzita, Přírodovědecká fakulta. 66 s..
- [14] TSYBAKOV, A. B. *Introduction to Nonparametric Estimation*. New York: Springer, 2009, 214 s.. ISBN 978-0-387-79051-0.
- [15] HÁTLE, J. a J. LIKEŠ. *Základy počtu pravděpodobnosti a matematické statistiky*. 2. vyd. Praha: SNTL, 1974, 464 s..
- [16] HÄRDLE, W. *Applied Nonparametric Regression*. Cambridge: Cambridge University Press, 1992, 409 s.. ISBN 978-0-521-42950-4.
- [17] TAKEZAWA, K. *Introduction to nonparametric regression*. 2. ed. Hoboken (New Jersey): Wiley, 2006, 538 s.. ISBN 0-471-74583-9.
- [18] ANDĚL, J. *Statistické metody*. 4. upr. vyd. Praha: Matfyzpress, 2007, 299 s.. ISBN 80-7378-003-8.
- [19] ANDĚL, J. *Základy matematické statistiky*. 2. opr. vyd. Praha: Matfyzpress, 2007, 358 s.. ISBN 80-7378-001-1.
- [20] EUBANK, R. L. *Nonparametric regression and spline smoothing*. 2nd ed. New York: Marcel Dekker, 1999, 337 s.. ISBN 0-8247-9337-4.
- [21] MAKRIDAKIS, S. G. S. C. WHEELWRIGHT a R. J. HYNDMAN. *Forecasting: methods and applications*. 3rd ed. Hoboken (New Jersey): Wiley, 1998. ISBN 0-471-53233-9.
- [22] CIPRA, T. *Finanční ekonometrie*. 2. upr. vy. Praha: Ekopress, 2013, 538 s.. ISBN 978-80-86929-93-4.
- [23] HOLLANDER, M. D. A. WOLFE a E. CHICKEN. *Nonparametric Statistical Methods*. 3rd ed. Hoboken (New Jersey): Wiley, 2014, 848 s.. ISBN 978-0-470-38737-5.
- [24] BIERENS, H. J. Kernel Estimators of Regression Functions, kap. 3. In: *Advances in Econometrics: Fifth World Congress*, sv. I. BEWLEY, T. F. ed. New York: Cambridge University Press, 1987, s. 99-144. ISBN 0521345529.
- [25] HALMOS, P. R. *Measure Theory*. New York: Springer, 1974, 304 s.. ISBN 0-387-90088-8. Graduate texts in Mathematics, 18.
- [26] BROM, P. *Předpověď kurzů akcií na krátké období*. Plzeň: 2013. Bakalářská práce. Západočeská univerzita v Plzni, Fakulta aplikovaných věd, Katedra matematiky. Vedoucí práce František Vávra. 51 s..
- [27] VÍŠKOVÁ, H. *Technická analýza akcií*. Praha: HZ Systém, 1997, 119 s.. ISBN 80-86009-13-0.

- [28] FIO BANKA, A.S. Ceník základních služeb. [online]. 23. 9. 2014 [cit. 2016-02-21]. Dostupné z: http://www.fio.cz/docs/cz/C_zaklad.pdf
- [29] FIO BANKA, A.S. Příloha ceníku základních služeb. [online]. 15. 1. 2016 [cit. 2016-02-21]. Dostupné z: http://www.fio.cz/docs/cz/C_trh.pdf
- [30] Fio banka. *Obchodování na pražské burze, BCPP* [online]. © 2016 [cit. 2016-03-24]. Dostupné z: <http://www.fio.cz/akcie-investice/obchodovani-akcie/akcie-cr/obchodovani-prazska-burza>
- [31] Burza cenných papírů Praha, a.s. *Seznam členů burzy* [online]. 1998-2016 [cit. 2016-Březen-18]. Dostupné z: <https://www.pse.cz/Clenove-Burzy/>
- [32] Klub investorů. *Jaký broker je ten pravý?* [online]. 8. 10. 2014 [cit. 2016-03-30]. Dostupné z: <http://www.klubinvestoru.com/cs/article/2285-jaky-broker-je-ten-pravy>
- [33] Daně při prodeji cenných papírů v 7 bodech. In: *Kurzycz* [online]. 28. 8. 2015, 9:26:02 SEČ [cit. 2016-02-21]. Dostupné z: <http://www.kurzycz.cz/zpravy/384848-dane-pri-prodeji-cennych-papiru-v-7-bodech/>
- [34] ČESKO. Zák. č. 586/1992 Sb. o daních z příjmů. [cit. 2016-02-21]. Dostupné z: <http://www.zakonyprolidi.cz/cs/1992-586>
- [35] ŘÍHA, J. *Technická analýza cenných papírů*. Praha: Comenia Nova, 1994, 103 s.. ISBN 80-901784-0-5.
- [36] Charles Dow - životopis. In: *Miras.cz* [online]. © 2000 - 2015 [cit. 2016-02-21]. Dostupné z: <http://www.miras.cz/akcie/investor-charles-dow.php>
- [37] SIEGEL, J. *Investice do akcií: běh na dlouhou trať*. Překlad Radomír ČÍŽEK. Praha: Grada, 2011, 296 s.. ISBN 978-80-247-3860-4.
- [38] CHEN, J. *Essentials of Technical Analysis for Financial Markets*. Hoboken (New Jersey): Wiley, 2010, 283 s.. ISBN 978-0-470-53729-9.
- [39] KIRKPATRICK, C. D. a J. DAHLQUIST. *Technical Analysis: The Complete Resource for Financial Market Technicians*. 2nd ed. Upper Saddle River (New Jersey): FT Press, 2010, 671 s.. ISBN 978-0-13-705944-7.
- [40] FANTA, J. *Počítačové analýzy kapitálových trhů*. KLENER, P. ed. Praha: Univerzita Karlova, 1999, 147 s.. ISBN 80-7184-895-6.
- [41] Akcie & dividendy. *Co jsou to dividendy* [online]. [cit. 2016-03-30]. Dostupné z: <http://www.akcieadividendy.cz/co-jsou-to-dividendy/>
- [42] Yahoo Help Central. *Finance | SLN2311 - About historical prices* [online]. [cit. 2016-03-30]. Dostupné z: <https://help.yahoo.com/kb/finance/SLN2311.html?impressions=true>

- [43] Sagefin. *Burzovní slovník* [online]. [cit. 2016-05-07]. Dostupné z: <http://www.sagefin.cz/burzovni-slovník/>
- [44] Skupina ČEZ. *Akcie* [online]. 2016 [cit. 2016-03-24]. Dostupné z: <http://www.cez.cz/cs/pro-investory/akcie.html>
- [45] Akcie.cz. *Kurzovní lístek akcie ČEZ* [online]. © 2013 [cit. 2016-05-05]. Dostupné z: <http://www.akcie.cz/kurzy-cz/historie/akcie-11392-cez/UTC=1463529600>
- [46] Patria.cz. *Komerční banka, a.s. - Oznámení Komerční banky o štěpení akcií* [online]. 25. 4. 2016 [cit. 2016-05-07]. Dostupné z: <http://www.patria.cz/zpravodajstvi/3189539/komercni-banka-as---oznameni-komercni-banky-o-stepeni-akcii.html>
- [47] Komerční banka. *Struktura akcionářů* [online]. © 2016 [cit. 2016-05-07]. Dostupné z: <https://www.kb.cz/cs/o-bance/vztahy-s-investory/akcionari/struktura-akcionarů/>
- [48] Philip Morris ČR a.s. (PM CR). *Informace pro akcionáře společnosti Philip Morris ČR a.s.* [online]. © 2002-2016 [cit. 2016-05-05]. Dostupné z: http://www.pmi.com/cs_cz/about_us/philip_morris_cr_shareholder_information/pages/philip_morris_cr_shareholder_information.aspx
- [49] Akcie.cz. *Kurzovní lístek akcie PHILIP MORRIS ČR* [online]. © 2013 [cit. 2016-05-05]. Dostupné z: <http://www.akcie.cz/kurzy-cz/historie/akcie-374-philip-morris-cr/>
- [50] Přehled platných smluv České republiky o zamezení dvojímu zdanění v oboru daní z příjmu. In: *Ministerstvo financí ČR* [online]. 2. 1. 2012, verze 10.2.2016 [cit. 2016-02-21]. Dostupné z: <http://www.mfcr.cz/cs/legislativa/dvoji-zdaneni/prehled-platnych-smluv>
- [51] ŽELEZNÝ, S. Jak zdanit příjmy ze zahraničních akcií a dividend? In: *Měšec.cz* [online]. 19. 8. 2013 [cit. 2016-02-22]. Dostupné z: <http://www.mesec.cz/clanky/jak-zdanit-prijmy-ze-zahranicnich-akcii-a-dividend/>
- [52] Technická analýza. In: *investice1.cz - Informační portál o investicích* [online]. 2013 - 2015 © [cit. 2016-02-21]. Dostupné z: <http://www.investice1.cz/technicka-analyza/>
- [53] FINANCIK.CZ. První krok k výdělkům: technické studium grafů. In: *Financnik.cz* [online]. 8. 12. 2004 [cit. 2016-02-21]. Dostupné z: <http://www.financnik.cz/komodity/manual/komodity-grafy.html>
- [54] Technická analýza (2.): Analýza grafických formací. In: *Miras.cz* [online]. © 2000 - 2015 [cit. 2016-02-21]. Dostupné z: <http://www.miras.cz/akcie/technicka-analyza-2.php>
- [55] BRO, R. et al. Creaminess of cream cheese: Multiway models for sensory profiling data. *Journal of Chemometrics* [online]. GEMPERLINE, P. J. ed. Wiley, 2008, **22** (1), 36-45 [cit. 2016-02-21]. ISRC 1099-128X. Data dostupná z: <http://www.models.life.ku.dk/Cream>. Dostupné z: <http://onlinelibrary.wiley.com/doi/10.1002/cem.1097/abstract>

- [56] HUBER, P. J. a E. RONCHETTI. *Robust statistics*. 2nd ed. Hoboken (New Jersey): Wiley, 2009, 354 s.. ISBN 978-0-470-12990-6.
- [57] WEISSTEIN, E. W. Multimodal. In: *MathWorld* [online]. © 1999-2016 [cit. 2016-29-02]. A Wolfram Web Resource. Dostupné z: <http://mathworld.wolfram.com/Multimodal.html>
- [58] WEISSTEIN, E. W. Disjoint sets. In: *MathWorld* [online]. © 1999-2016 [cit. 2016-03-04]. A Wolfram Web Resource. Dostupné z: <http://mathworld.wolfram.com/DisjointSets.html>
- [59] Česká národní banka. *Slovník pojmů* [online]. © 2003-2016 [cit. 2016-03-24]. Dostupné z: <http://www.cnb.cz/cs/obecne/slovník/index.html>
- [60] Patria Direct. *Slovník investora* [online]. [cit. 2016-03-24]. Dostupné z: <https://www.patria-direct.cz/vzdelavani/zaciname-s-investovanim/slovník-investora>
- [61] Zlatá koruna. *Porovnávač produktů - Obchodování s CP* [online]. © 2003 - 2016 [cit. 2016-03-30]. Dostupné z: <http://www.zlatakoruna.info/financni-produkty/obchodovani-s-cp>
- [62] BUDÍKOVÁ, M. T. LERCH a Š. MIKOLÁŠ. *Základní statistické metody*. Brno: Masarykova univerzita, 2009, 180 s.. ISBN 978-80-210-3886-8.
- [63] MCLACHLAN, G. J. a D. PEEL. *Finite Mixture Models*. New York: Wiley, 2000, 456 s.. ISBN 0-471-00626-2.
- [64] O2 Czech Republic. *Vztahy s investory* [online]. [cit. 2016-05-07]. Dostupné z: <https://www.o2.cz/spolecnost/vztahy-s-investory/>

Příloha A – přiložené soubory

Seznam skriptů přiložených na CD:

- 1) Skripty určené pro proces optimalizace popsany v podkapitole 6.2 a v části 6.4.1
 - a. *Optimalizace.m* ... hlavní (spouštěcí) program optimalizační části práce. „Optimální“ nastavení pro zvolený akciový titul se vždy vypíše v konzoli. Podpůrné skripty:
 - *generuj_signaly.m*
 - *nastaveni_optimalizace.m*
 - *odhadni_funkci.m*
 - *vyfiltruj_signaly.m*
 - b. *nastaveni_optimalizace.m* ... skript, který vytvoří uživatelské menu pro volbu nastavení optimalizace (viz Obrázek 6.10).
 - c. *vyfiltruj_signaly.m* ... eliminuje nadbytečné signály tak, aby vždy jeden prodejní signál následoval jeden nákupní signál (viz část 6.2.3, funkce (48))
- 2) Skripty určené pro simulaci obchodního systému především v podkapitole 6.3 a v části 6.4.2
 - a. *Obchodni_system.m* ... obchodní systém pro obchodování s akcemi na BCPP prostřednictvím internetové aplikace e-broker společnosti Fio (podle toho jsou také započítávány poplatky) (pro popis algoritmu viz část 6.3.2). Podpůrné skripty:
 - *generuj_signaly.m*
 - *nastaveni_simulace.m*
 - *odhadni_funkci.m*
 - *vypocitej_poplatek.m*
 - b. *nastaveni_simulace.m* ... skript, který vytvoří uživatelské menu pro volbu nastavení simulace obchodního systému (viz Obrázek 6.6).
 - c. *vypocitej_poplatek.m* ... vypočítá velikost poplatku (v poměru k objemu obchodu) za obchod, tj. náklady související s aktuálním příkazem (založeno na Tabulka 2)
- 3) Skripty používané oběma předchozími skupinami
 - a. *generuj_signaly.m* ... generování signálů k nákupu a prodeji akcií podle pravidel použití více odhadů používaných technickou analýzou, možnost použití jako funkční obchodní systém (viz část 6.2.3, funkce (47)). Podpůrný skript:
 - *vypocitej_poplatek.m*
 - b. *odhadni_funkci.m* ... odhad funkce hustoty nebo regresní křivky. Možnost odhadu pomocí klouzavých průměrů, klouzavé regrese nebo jádrového odhadu. (viz část 6.2.3, funkce (46)).

Seznam vstupních a výstupních souborů přiložených na CD:

- 1) *Data.xlsx* ... soubor se zdrojovými vstupními daty, včetně provedených úprav o dividendy
- 2) *Optimalizace-vystup.xlsx* ... soubor s výsledky každé iterace procesu optimalizace podle části 6.2.1 i podkapitoly 6.4

Příloha B – zdrojové kódy

Zdrojový kód skriptu pro odhad tvaru funkce - *odhadni_funkci.m*:

```
1 function
2 [odhad]=odhadni_funkci(rezim,data,x,h,druhOdhadu,zpozdeny,zpozdeni)
3
4 % [odhad]=odhadni_funkci(rezim,data,x,h,druhOdhadu,zpozdeny,zpozdeni)
5 %
6 % Funkce provadejici odhad funkce hustoty nebo regresni krivky. Moznost
7 % odhadu pomoci klouzavych prumeru, klouzave regrese nebo jadroveho
8 % odhadu.
9 %
10 % Predpoklada se, ze vstupy jsou sloupceve vektory.
11 %
12 % Vstupy:
13 % rezim      ... rezim spusteni funkce - bud „hustota“ nebo „regrese“
14 % data       ... odhadovana data
15 % x         ... hodnoty, ve kterych ma byt odhad proveden
16 % h         ... vyhlazovací parametr nebo delka klouzaveho okna
17 % druhOdhadu ... jadrova funkce nebo klouzavy prumer, který bude k odhadu
18 %            pouzit
19 %            - bud „s“, „w“, „x“, „r“, „p“, „e“, „t“ nebo „g“
20 % zpozdeny   ... promenna urcujici, zda pujde o „zpozdeny“ jadrový odhad
21 % zpozdeni   ... parametr „posunu“ jadroveho odhadu
22 %
23 % Vystupy:
24 % odhad ... odhad funkce
25 %          - funkce hustoty
26 %          - regresni krivka (Nadaraya-Watsonuv vzorec, klouzave
27 %          prumery)
28 %
29 % Soucast diplomove prace
30 % autor: Bc. Pavel Brom
31 % osobni cislo: A15N0008P
32 %% Osetreni vyjimek
33 if strcmp(rezim,'hustota') && (~strcmp(druhOdhadu,'p') &&...
34     ~strcmp(druhOdhadu,'e') && ~strcmp(druhOdhadu,'t') &&...
35     ~strcmp(druhOdhadu,'g'))
36     error('K odhadu funkce hustoty lze pouzit pouze jednu z definovanych
37     jadrovych funkci!')
38 end
39
40 if nargin<6
41     zpozdeny=false;
42 end
43
44 if ~zpozdeny && strcmp(druhOdhadu,'s') && mod(h,2)==0
45     error('Pro nezpozdeny jednoduchy klouzavy prumer je nutne volit liche
46     h!')
47 end
48
49 if strcmp(druhOdhadu,'w') && ~zpozdeny
50     % vzdy se chova jako zpozdeny
51     warning('Vazeny klouzavy prumer se vzdy chova jako "zpozdeny"!')
52     zpozdeny=true;
53 end
```

```

54 if strcmp(druhOdhadu,'x') && ~zpozdeny
55     % vzdy se chova jako zpozdeny
56     warning('Exponencialni klouzavy prumer se vzdy chova jako "zpozdeny"!
57 Jde o vynechani prvnich (h-1) hodnot.')
58     zpozdeny=true;
59 end
60
61 if strcmp(druhOdhadu,'r') && ~zpozdeny
62     % vzdy se chova jako zpozdeny
63     warning('Klouzava regrese se vzdy chova jako "zpozdena"!')
64     zpozdeny=true;
65 end
66
67 if zpozdeny
68     % zpozdeny odhad
69     if nargin==6
70         zpozdeni=h-1;
71     else
72         zpozdeni=zpozdeni-1;
73     end
74 else
75     zpozdeni=0;
76 end
77 odhad=nan(size(x,1),size(x,2));
78 %% Odhad funkce
79
80 if strcmp(druhOdhadu,'s')
81     % Jednoduchy klouzavy prumer
82     if ~zpozdeny
83         % klasicky odhad regresni krivky
84         p=(h-1)/2;
85         n=length(data);
86         for i=1:n
87             Y=[data(p:-1:1),data,data(n-p+1:end)];
88             j=1-p:n+p;
89             odhad(i)=sum(Y(-p<=(i-j) & (i-j)<=p)/h);
90         end
91     else
92         % zpozdeny odhad
93         for i=1+zpozdeni:length(data)
94             odhad(i)=mean(data(i-zpozdeni:i));
95         end
96     end
97 elseif strcmp(druhOdhadu,'w')
98     % Vazeny klouzavy prumer
99     vahy=2/(h*(h+1));
100    for i=1+zpozdeni:length(data)
101        pom=0;
102        for j=0:h-1
103            pom=pom+(h-j)*data(i-j);
104        end
105        odhad(i)=vahy*pom;
106    end
107 elseif strcmp(druhOdhadu,'x')
108     % Exponencialni klouzavy prumer
109     alfa=2/(h+1);
110    for i=1+zpozdeni:length(data)
111        pom=0;
112        for j=0:i-2
113            pom=pom+alfa*(1-alfa)^j*data(i-j);
114        end
115        odhad(i)=pom+(1-alfa)^(i-1)*data(1);
116    end

```

```

117 elseif strcmp(druhOdhadu, 'r')
118     % Klouzava regrese
119     X=[ones(h,1) (1:h)'];
120     for i=1+zpozdeni:length(data)
121         b=(X'*X)\X'*data(i-zpozdeni:i);
122         odhad(i)=b(1)+b(2)*(h+1);
123     end
124 else
125     % Jadrovy odhad
126
127     % Jadrove funkce
128     if strcmp(druhOdhadu, 'p')
129         konstanta=sqrt(3);
130         jadro=@(x) repmat(1/(2*konstanta),size(x,1),size(x,2));
131     elseif strcmp(druhOdhadu, 'e')
132         konstanta=sqrt(5);
133         jadro=@(x) 3*(1-(x./konstanta).^2)/(4*konstanta);
134     elseif strcmp(druhOdhadu, 't')
135         konstanta=sqrt(6);
136         jadro=@(x) (1-abs(x./konstanta))./konstanta;
137     elseif strcmp(druhOdhadu, 'g')
138         konstanta=inf;
139         jadro=@(x) exp(-x.^2/2)/sqrt(2*pi);
140     end
141
142     %% Jadrovy odhad
143     for j=1+zpozdeni:length(x)
144         if strcmp(rezim, 'hustota')
145             % "xx" = "sort(data)"
146             argument=(x(j-zpozdeni)-data)/h;
147         elseif strcmp(rezim, 'regrese')
148             % v bodech pozorovani
149             argument=(x(j-zpozdeni)-x)/h;
150         end
151         kx=jadro(argument);
152         kx(argument<=-konstanta | argument>konstanta)=0;
153
154         if strcmp(rezim, 'hustota')
155             % jadrovy odhad funkce hustoty
156             odhad(j)=sum(kx)/(length(data)*h);
157         elseif strcmp(rezim, 'regrese')
158             % Nadaraya-Watsonuv jadrovy odhad regresni krivky
159             if sum(kx)~=0
160                 odhad(j)=sum(kx.*data)/sum(kx);
161             else
162                 odhad(j)=0;
163             end
164         end
165     end
166 end
167
168 end

```

Zdrojový kód skriptu klíčového pro navržený obchodní systém - *generuj_signaly.m*:

```
1 function
2 [nakup, prodej, nakupVarovani, prodejVarovani, zisky, poplatky, stav_uctu]=gene
3 ruj_signaly(y1,y2,y3,h,rezim,stav_uctu,kurzy)
4
5 % [nakup,prodej,nakupVarovani,prodejVarovani,zisky,poplatky,stav_uctu]=
6 % generuj_signaly(y1,y2,y3,h,rezim,stav_uctu,kurzy)
7 %
8 % Funkce pro generovani signalu k nakupu a prodeji akcií podle pravidel
9 % pouziti více odhadu pouzivanych technickou analyzou, možnost pouziti jako
10 % funkcní obchodní systém
11 %
12 % Vstupní parametry:
13 % y1      ... buď kurz nebo odhad kratkodobeho charakteru
14 % y2      ... buď odhad kratkodobeho (v případě, že y1 je kurz), odhad
15 %          Strednedobeho (v případě srovnávání 3 odhadu) nebo
16 %          dlouhodobého charakteru (v případě, že y1 je odhad
17 %          kratkodobeho charakteru)
18 % y3      ... odhad dlouhodobého charakteru (pokud není k dispozici
19 %          vyplnit prázdné pole [])
20 % h      ... parametr posunu y2 nebo y3
21 % rezim   ... režim spuštění funkce, tj. buď pouze generování signalu
22 %          nebo obchodní systém, tj. reakce na signály
23 % stav_uctu ... aktuální stav obchodního účtu
24 % kurzy   ... kurzy akcií
25 %
26 % Vystupní parametry:
27 % nakup   ... vygenerované signály k nakupu
28 % prodej  ... vygenerované signály k nakupu
29 % nakupVarovani ... vygenerované varovné signály k nakupu
30 % prodejVarovani ... vygenerované varovné signály k prodeji
31 % zisky   ... zisky z jednotlivých uzavřených obchodů (vystup pouze
32 %          v režimu „obchod“)
33 % poplatky ... poplatky za realizaci jednotlivých signalu na trhu
34 %          (vystup pouze v režimu „obchod“)
35 % stav_uctu ... stav obchodního účtu po provedení všech vygenerovaných
36 %          příkazů k nakupu a prodeji (vystup pouze v režimu
37 %          „obchod“)
38 %
39 % Pro pouziti je nutná podpurná funkce:
40 %     vypocitej_poplatek.m
41 %
42 % Součást diplomové práce
43 % autor: Bc. Pavel Brom
44 % osobní číslo: A15N0008P
45
46 if nargin<4
47     error('Nedostatek vstupnich parametru! -> chybi vyhlazovaci
48 parametr!')
49 elseif nargin==4
50     % implicitně pouze generuje signály
51     rezim='signaly';
52 elseif nargin==5
53     error('Nedostatek vstupnich parametru! -> chybi aktualni stav uctu!')
54
55 elseif nargin==6
56     error('Nedostatek vstupnich parametru! -> chybi vyvoj kurzu akcií!')
57 end
58
59 if ~strcmp(rezim,'signaly') && ~strcmp(rezim,'obchod')
60     error('Spatne zvoleny rezim! -> buď "signaly" nebo "obchod"!')
61 end
```

```

62  if length(y1)<h
63      error('Nedostatek pozorovani pro aktualne zadane parametry!!!')
64  end
65
66  if strcmp(rezim,'obchod')
67      dlouha=false; % indikator aktualni pozice
68      zisky=[]; % zisky/ztraty za kazdy relizovany (uzavreny) obchod
69      poplatky=[]; % poplatky za kazdy potvrzeny prikaz
70  end
71
72  if isempty(y3)
73      % vstup: (y1,y2,[],h)
74      % bud kurz a odhad nebo 2 odhady s ruznymi delkami okna
75
76      % 2 prumery
77      nakup=false(h-1,1);
78      nakup=[nakup;false]; % signal k nakupu mohu dostat, az kdyz znam kurz
79      % v case po prekrizeni
80      prodej=false(h-1,1);
81      prodej=[prodej;false];
82      for i=h+1:length(y1)
83          if y1(i-1)<=y2(i-1) && y1(i)>y2(i)
84              prodej(i)=false;
85
86              if strcmp(rezim,'signaly')
87                  nakup(i)=true;
88              else
89                  nakup(i)=false;
90                  if ~dlouha
91                      % zatim v zadne pozici -> lze nakoupit -> dlouha pozice
92                      dlouha=true;
93                      nakup(i)=true;
94                      % maximalizace investovane castky
95                      pocet=floor(stav_uctu/kurzy(i)); % pocet nakoupenych
96                      % akci v case t
97                      funObj=@(ks)ks*kurzy(i)* ...
98                          (1+vypocitej_poplatek(ks*kurzy(i)));
99                      funcon=@(ks)deal(stav_uctu-ks*kurzy(i)*...
100                          (1+vypocitej_poplatek(ks*kurzy(i))),[]);
101                      x0=pocet;
102                      options = optimset('Algorithm','sqp');
103                      [poc,fval] = fmincon(funObj,x0,[],[],[],[],[],[],...
104                          funcon,options);
105                      pocet=floor(poc);
106                      objem=pocet*kurzy(i); % objem obchodu pri nakupu akci
107                      [zbytek,~]=funcon(pocet); % neinvestovana castka
108                      poplatek=vypocitej_poplatek(objem);
109                      ucetPuv=stav_uctu;
110                      stav_uctu=stav_uctu-objem*(1+poplatek);
111
112                      poplatky=[poplatky objem*poplatek];
113                  else
114                      % jiz v dlouhe pozici ale opet signal k nakupu
115                  end
116              end
117          elseif y1(i-1)>=y2(i-1) && y1(i)<y2(i)
118              nakup(i)=false;
119
120              if strcmp(rezim,'signaly')
121                  prodej(i)=true;
122              else
123                  prodej(i)=false;
124                  if dlouha

```

```

125         % jiz v dlouhe pozici -> lze prodat
126         dlouha=false; % vystoupeni z dlouhe pozice
127         prodej(i)=true;
128         objem=pocet*kurzy(i);
129         poplatek=vypocitej_poplatek(objem);
130         stav_uctu=stav_uctu+objem*(1-poplatek);
131         zisky=[zisky stav_uctu-ucetPuv];
132         poplatky=[poplatky objem*poplatek];
133     else
134         % nejsme v dlouhe pozici tak ani neni, co prodat
135         % tato situace pri definovanych podminkach nenastane
136     end
137 end
138 else
139     nakup(i)=false;
140     prodej(i)=false;
141 end
142 end
143 nakupVarovani=[];
144 prodejVarovani=[];
145 else
146     % vstup: (y1,y2,y3,h)
147     % 3 odhady s ruznymi delkami okna
148     nakupVarovani=false(h-1,1);
149     nakup=false(h-1,1);
150     nakupVarovani=[nakupVarovani;false]; % signal k nakupu mohu dostat,
151     % az kdyz znam kurz v case po prekrizeni
152     nakup=[nakup;false];
153     prodejVarovani=false(h-1,1);
154     prodej=false(h-1,1);
155     prodejVarovani=[prodejVarovani;false];
156     i=h+1;
157     % generovani signalu
158     while i<=length(y1)
159         if y1(i-1)<=y3(i-1) && y1(i)>y3(i)
160             % varovani
161             nakupVarovani(i)=true;
162             prodejVarovani(i)=false;
163             while i<=length(y1) && y1(i)>y3(i)
164                 if (y2(i-1)<=y3(i-1) && y2(i)>y3(i)) || y2(i)>y3(i)
165                     % potvrzeni
166                     % strednedoby bud se taky prave prekrizil nebo protnul
167                     % uz drive
168                     if i>length(nakupVarovani)
169                         nakupVarovani(i)=false;
170                         prodejVarovani(i)=false;
171                     end
172                 end
173                 prodej(i)=false;
174             end
175             if strcmp(rezim,'signaly')
176                 nakup(i)=true;
177             else
178                 nakup(i)=false;
179                 if ~dlouha
180                     % zatim v zadne pozici -> lze nakoupit ->
181                     % dlouha pozice
182                     dlouha=true;
183                     nakup(i)=true;
184                     % maximalizace investovane castky
185                     % pocet nakoupenych akci v case t
186                     pocet=floor(stav_uctu/kurzy(i));
187                     funObj=@(ks)ks*kurzy(i)*...
188                         (1+vypocitej_poplatek(ks*kurzy(i)));

```

```

189         funcon=@(ks)deal(stav_uctu-ks*kurzy(i)*...
190             (1+vyprocitej_poplatek(ks*...
191                 kurzy(i))),[]);
192     x0=pocet;
193     options = optimset('Algorithm','sqp');
194     [poc,fval] = fmincon(funObj,x0,[],[],[],[],...
195         [],[],[],funcon,options);
196     pocet=floor(poc);
197     % objem obchodu pri nakupu akcii
198     objem=pocet*kurzy(i);
199     % neinvestovana castka
200     [zbytek,~]=funcon(pocet);
201     poplatek=vyprocitej_poplatek(objem);
202     ucetPuv=stav_uctu;
203     stav_uctu=stav_uctu-objem*(1+poplatek);
204
205     poplatky=[poplatky objem*poplatek];
206     else
207         % jiz v dlouhe pozici ale opet signal k nakupu
208     end
209     end
210
211     i=i+1;
212     break;
213     end
214     if i>length(nakupVarovani)
215         nakupVarovani(i)=false;
216         prodejVarovani(i)=false;
217     end
218     nakup(i)=false;
219     prodej(i)=false;
220     i=i+1;
221     end
222 elseif y1(i-1)>=y3(i-1) && y1(i)<y3(i)
223     % varovani
224     nakupVarovani(i)=false;
225     prodejVarovani(i)=true;
226     while i<=length(y1) && y1(i)<y3(i)
227         if (y2(i-1)>=y3(i-1) && y2(i)<y3(i)) || y2(i)<y3(i)
228             % potvrzeni
229             % strednedoby bud se taky prave prekrzil nebo protnul
230             % uz drive
231             if i>length(nakupVarovani)
232                 nakupVarovani(i)=false;
233                 prodejVarovani(i)=false;
234             end
235             nakup(i)=false;
236             if strcmp(rezim,'signaly')
237                 prodej(i)=true;
238             else
239                 prodej(i)=false;
240                 if dlouha
241                     % jiz v dlouhe pozici -> lze prodat
242                     dlouha=false; % vystoupeni z dlouhe pozice
243                     prodej(i)=true;
244                     poplatek=vyprocitej_poplatek(objem);
245                     stav_uctu=stav_uctu+pocet*kurzy(i)*...
246                         (1-poplatek);
247                     zisky=[zisky stav_uctu-ucetPuv];
248                     poplatky=[poplatky pocet*kurzy(i)*poplatek];
249                 else
250                     % nejsme v dlouhe pozici tak ani neni, co
251                     % prodat

```

```

252                                     % tato situace pri definovanych podminkach
253                                     % nenastane
254                                     end
255                                     end
256                                     i=i+1;
257                                     break;
258                                     end
259                                     if i>length(nakupVarovani)
260                                         nakupVarovani(i)=false;
261                                         prodejVarovani(i)=false;
262                                     end
263                                     nakup(i)=false;
264                                     prodej(i)=false;
265                                     i=i+1;
266                                     end
267                                     else
268                                         nakupVarovani(i)=false;
269                                         prodejVarovani(i)=false;
270                                         nakup(i)=false;
271                                         prodej(i)=false;
272                                         i=i+1;
273                                     end
274                                     end
275                                     end
276
277                                     end

```